# University of Naples Federico II

Department of Agricultural Science

PhD course in Agricultural and Agri-Food Sciences

XXXII Cycle (2017-2020)

# Selection of superior heat tolerant tomato lines through genetic and genomic strategies

PhD candidate:

Dott. Elisa Cappetta

Supervisor:

Prof. Maria Raffaella Ercolano

*A mia madre*

# CONTENT

## ABSTRACT

Climate is unequivocally changing. The planet's average surface temperature has risen about 1.62 degrees Fahrenheit (0.9 degrees Celsius) since the late 19th century. Most of the warming climates occurred in the past 35 years, were recorded in the last 10 years. Global warming affects agriculture in a number of ways, including temperate region average temperature changes and climate extremes. Tomato (*Solanum lycopersicum*) is both an important commercial crop and a model system for genetic studies, due to its diploid, relatively compact, and recently sequenced genome and to the availability of large genetic and genomic resource collections. Tomato species is highly sensitive to high temperature and few degrees above its optimum growth temperature threshold can lead to serious deleterious effects, such as flower abscission, decrease of pollen quality, abnormal growth, reduced fruit set and yield. Therefore, the development of innovative strategies to obtain tomato cultivars with improved yield under high temperature conditions is a main goal for plant molecular science and breeding.

In this thesis, different breeding strategies were used to improve tomato cultivars tolerance to heat stress. A F4 segregating population, deriving from the tomato variety JAG8810, previously selected for yield performance under hot stress, was phenotypically investigated under heat stress conditions to evaluate quantitative and qualitative traits. By the means of a PCA analysis, best and worst performers were selected. Extreme individuals for yield were also evaluated for sub-traits, such as pollen viability, to better investigate the basis of heat stress tolerance and correlations among heat tolerance component traits. In addition, the cultivar *Moneymaker* was tested for a heat treatment on limited areas, suggesting that heat stress response is a local process. The F4 population deriving from the JAG8810 variety, was sequenced by genotyping by sequencing (GBS) approach to identify all possible variants. Genomic prediction models for yield production per plant (YP) and soluble solid content (SSC) under heat stress, were developed. Several parameters, including training population size and composition and marker quality were adjusted to obtain optimized models for assessed traits and population. The predicted GEBVs (genetic breeding values) of F5 offspring were phenotypically validated in field. Furthermore, the most meaningful SNPs selected for model construction were used to conduct a QTL analysis to shed light on the genetic basis of heat tolerant traits in tomato. The analysis permitted the identification of 5 QTLs involved in yield and one in SSC. Two candidate genes putatively involved in heat stress tolerance were discovered in regions underlining QTLs. Finally, with the aim to identify regulatory elements involved in the abiotic stress tolerance, a tomato genome scan and a phylogenetic analysis of Dof proteins was performed identifying SlDof11 as suitable target for CRISPR/Cas9 experiments.

# CHAPTER I

## GENERAL INTRODUCTION

### 1.1 HEAT STRESS IN PLANTS: A BRIEF OVERVIEW

Climate is unequivocally changing. The planet's average surface temperature has risen about 1.62 degrees Fahrenheit (0.9 degrees Celsius) since the late 19th century, a change mainly driven by atmospheric greenhouse gases increase, wide-spread land transformation and other human-made emissions into the atmosphere (Asseng et al., 2015). Most of the warming occurred in the past 35 years, with the five warmest years on record taking place since 2010 (figure 1) (https://climate.nasa.gov/).

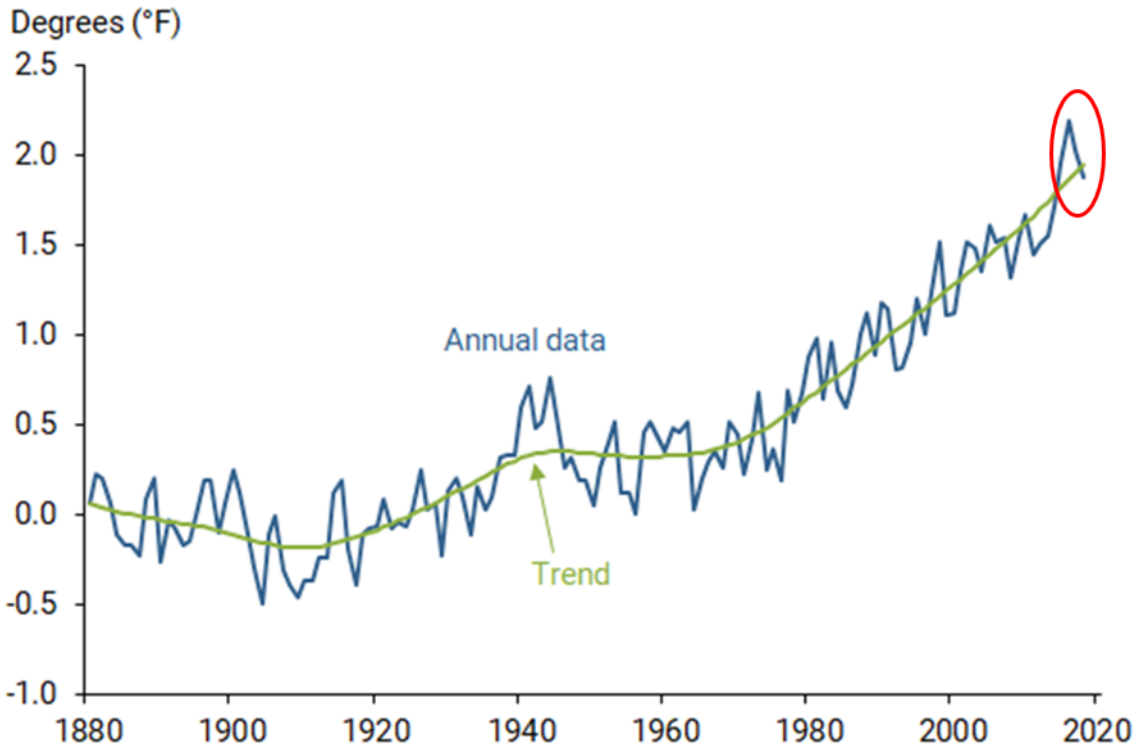



**Figure 1. Change in global average temperature since 1880. The red circle highlights the increase in temperature in the last five years.**

The "global warming" process has received increasing attention in recent years, due to its far-reaching effects on human society, including the extinction of species that cannot escape their environment and a substantial reduction in crop productivity (Wheeler and von Braun, 2013; Fahad et al., 2017).

Indeed, each plant species has an optimal temperature range for growth, and each increase above the optimal temperature range can be considered as heat stress. Serious losses in crop production may occur because many biological developmental processes are sensitive to high temperature and they may be irreversibly damaged. Therefore, the scientific community is improving efforts for deepening our knowledge about Heat Stress Response (HSR) in plants.

The effects of high-temperature can be direct if they induce modification in existing physiological processes, or indirect if they alter patterns of development. Moreover, adverse effects of high temperature vary with the intensity, duration and sternness of the stress (Morimoto R., 1998; Wahid A et al., 2007; Fahad et al., 2016). Examples of direct injuries are protein denaturation and aggregation, and increased fluidity of membrane lipids. Examples of indirect damages are inactivation of enzymes in chloroplasts and mitochondria (affecting photosynthesis and cell respiration), inhibition of protein synthesis, protein degradation, loss of membrane integrity and alteration of the correct organization of microtubules (Wahid et al., 2007). In addition to these damages, heat stress may induce oxidative damage increasing production and accumulation of Reactive Oxygen Species (ROS) that rapidly become excessive (Volkov et al., 2006; Driedonks et al., 2015). This ROS accumulation pose serious threat to the cell functioning by damaging lipids and proteins (Apel and Hirt, 2004; Bita et al., 2011; Fahad et al., 2017).

Immediately after exposure to high temperatures, plants make many changes at the molecular level, including the expression of genes and accumulation of transcripts. This can lead to the synthesis of stress-related proteins, called heat shock proteins (HSPs), as a key component of a stress-tolerance strategy (Iba, 2002; Wahid et al., 2007; Hanjing, 2015).

A short exposure to very high temperature or a long period of exposure may cause severe cellular damages and even cell death as a consequence of the collapse of cellular organization. Indeed, injuries can be too severe to be healed by stress induced- protection pathways. Furthermore, plants respond differentially to high temperatures depending on species, genotype, organ, or developmental stage (Nguyen et al., 2019). In particular, the reproductive phase is more sensitive to high temperature compared to the vegetative growth phase (Warland et al., 2006; Hasanuzzaman et al., 2013; Ruggieri et al., 2019). Besides, the effect of the heat on the organism can be local, when heat effect is only registered in the part of the plant directly exposed to heat, or systemic, if after exposure of only a portion of the plant, the HSR appears in the entire organism. All these factors make the HSR a complex system to investigate.

### 1.1.2 EFFECTS OF HEAT STRESS ON REPRODUCTIVE SYSTEM

An increasing in temperature can damage every plant tissue or organ. Numerous studies document a reduction in final yields due to heat stress during flowering since the reproductive apparatus resulted to be very sensitive in many crops like rice, wheat, barley and tomato (Barnabàset al., 2006; Zinn et al., 2010; Jagadish et al., 2014; Driedonks et al., 2016; Xu et al., 2017).

The heat stress effect on reproductive system acts on multiple levels. Generally, a moderate heat stress will often accelerate flowering, which may cause reproduction to occur before plants accumulate adequate resources for allocation to developing seeds (Balasubramanian et al., 2006; Tonsor et al., 2008). Heat stress can have a different effect on male and female structures and, as a consequence, it may create a lack of synchrony between the development of the two tissues (Herrero, 2003; Hedhly et al., 2008). A third category of heat stress effects includes defects in the structure and function of parental tissues (i.e. corollas and stamens) reducing the number, decrease the size, and cause deformity of floral organs (Takeoka et al., 1991; Morrison and Stewart, 2002). Finally, although both reproductive organs are affected, little is known on the effects of heat stress on female reproductive organs. Whereas, the effects of temperature stress on male reproductive structures has been widely studied by highlighting that pollen development is the most heat-sensitive process in plant reproduction (Sato, 2002; Barnabaset et al., 2008). Indeed, the sudden decline in yield due to high temperatures experienced by plants during flowering is mainly associated with pollen infertility (Zinn et al., 2010; Bita et al., 2013).

Reduction in pollen viability after exposure to heat has been observed in several crop species like barley (Sakata et al., 2010), tomato (Pressman et al., 2002; Firon et al., 2012; Müller et al., 2016), soybean (Djanaguiraman et al., 2013) and rice (Matsui and Omasa, 2002). It has been established that a heat stress at early stages of pollen grains development may cause an arrest in the biological process (De Storme and Geelen, 2013). The developmental sensitivity to heat stress in pollen seems to be species-specific, even if meiosis is likely to be the most sensitive process for many crops (Paupière et al., 2014). Unfortunately, reduced fertility at high temperature cannot be clearly identified as a defect in a single function, but most likely varies with the type of heat stress and the plant species. Loss of fertility could result from problems in male meiosis, pollen germination, pollen tube growth, or megagametophyte defects among other factors. Recently, it has been suggested that a reduction in pollen number and viability might be the indirect result of defects in the tapetal cells (De Storme and Geelen, 2014; Xu et al., 2017). For example, in wheat, heat stress during the period of microspore meiosis can induce tapetum degradation (Sakata et al., 2010) that leads to pollen sterility. Similar tapetal defects, always associated with reduced pollen viability, are known from cold and drought

stress, and occur in different plant species such as wheat, barley, rice, Arabidopsis and tomato (De Storme and Geelen, 2014; Parish et al., 2012).

## 1.2 BREEDING METHODS FOR OBTAINING HEAT STRESS TOLERANT PLANTS

The traditional method for breeding for heat tolerance is to grow advanced lines in a hot target production environments and select lines that have greater yields than current cultivars (this provides a direct measure of heat tolerance). However, high-temperature tolerance is a complex phenomenon highly influenced by environmental variations, controlled by multiple genes imparting a number of physiological and biochemical changes. Due to uncontrolled environmental factors and the influence of additional abiotic and biotic stresses, it can be difficult to select for high-temperature tolerance through conventional breeding (Tayade et al., 2018). Suitable breeding techniques to improve crop species in order to easy produce elite lines with heat tolerance traits are therefore surely needed to address this challenge (Driedonks et al., 2016).

In recent years, the development of molecular-marker techniques and their applications drastically changed the fate of plant breeding. Molecular markers were mainly integrated in traditional phenotypic selection (PS) by applying marker-assisted selection (MAS) to improve the plant selection process through the selection of chromosomal segments containing quantitative trait locus (QTL).

Few genetic selection approaches mediated by the aid of molecular markers on heat tolerance in different crops have been conducted (Cao et al., 2003; Zhao et al., 2006). In some plants, such as Arabidopsis, barley, brassica, cowpea, maize, potato, rice, sorghum, tomato and wheat, QTLs for heat tolerance-related traits have been discovered, including reproductive traits such as yield, fruit set, grain weight, grain filling rate, days to heading, spikelet fertility, pollen germinability and pollen tube growth (Jha et al., 2014).

Traits, such as yield, are often controlled by many “minor” gene effects, which are difficult to map. However, also if mapping is successful, often multiple quantitative trait loci (QTL) are present, which are difficult to locate and as a consequence, the proportion of genetic variance that is capitalized by MAS is limited (Goddard and Hayes, 2009). For traits with polygenic inheritance, a strategy able to predict the genomic potential of any individual should be used. A GS approach usually allows a better prediction of a genotype potential performance since it provides a direct estimation of the likelihood that it possesses superior alleles (Bassi et al., 2016). In this regard, genomic selection (GS) provides new opportunities for increasing the efficiency of plant breeding programs (Bernardo and Yu, 2007; Heffner et al., 2009; Crossa et al., 2010; Lorenz et al., 2011) by reducing the cost per cycle and the time required for developing a new variety (figure 2).

The aim of GS is to estimate the genetic potential of the candidates for selection using genome-wide marker data. Basically, GS combines genotypic and phenotypic data in a training population (TRN) to obtain the genomic estimated breeding values (GEBVs) of individuals in a testing population (TST) that have been genotyped but not phenotyped.

The decreasing costs of high-density genotyping system based on single nucleotide polymorphism (SNP) and development of statistical methods capable of accurately predict marker effects have led to the breakthrough of GS increasing the rate of genetic gain per unit of time. Nowadays, GS has been used in several crops (Lorenz et al., 2011; Crossa et al., 2014; Song et al., 2017; Yamamoto et al., 2017; Cui et al., 2020) and empirical studies have demonstrated that GS has higher genetic gain than MAS for traits controlled by large number of QTL (Crossa et al., 2017).

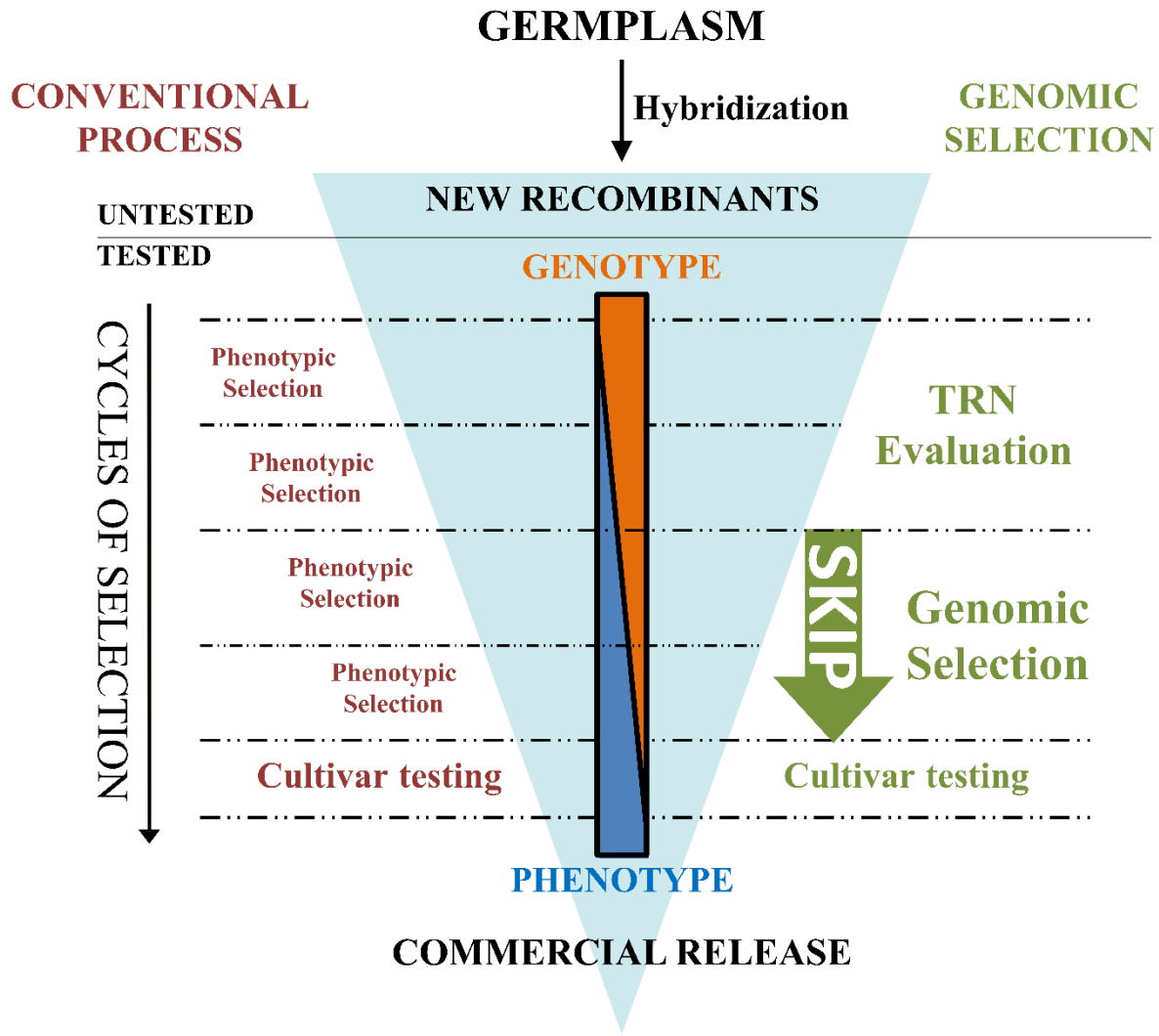



**Figure 2. Comparison of genomic selection and conventional process for plant breeding. Genomic screening of germplasm accelerates and optimizes the genetic gain obtained in each selection cycle (from Cappetta et al 2020).**

## 1.3 BIOTHECNOLOGICAL APPROACHES TO IMPROVE HEAT STRESS TOLERANCE

Breeding systems described above allowed breeders to produce improved varieties of many crops. However, many challenges persist and alternative innovative technologies have the potential to address many of these challenges. For example, in the last decade, new technologies, commonly referred to as genome-editing technologies, have revolutionized the field of plant science and agriculture. These technologies, that rely on engineered site-specific nucleases (SSNs), can overcome incompatibility barriers between different species and introduce artificial/synthetic genes into crop plants; and hence, generate new cultivars with novel traits. In contrast to the transgenic approach, which leads to random insertions and very often random phenotypes, genome editing methods produce defined mutants, thus becoming a potent tool in functional genomics and crop breeding (Malzahn et al., 2017). Improved crops can be used in breeding programs and the resulting varieties can be directly used (Waltz, 2018). The SSNs such as zinc finger nucleases (ZFNs) and transcriptional activator-like effector nucleases (TALENs) operate through the fusion of sequence-specific DNA binding domains (DBDs) and nucleases. Following the recognition of the target sequence by the DBDs, nucleases provide double-strand breaks (DSBs) leading to loss-of-function mutations (Chandrasegaran and Carroll, 2016). More recently, the CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats) /Cas9 (CRISPR-associated protein-9 nuclease) technology was developed consisting of engineered nucleases driven to the target sequence by a specifically designed guide RNA (gRNA) (figure 3).

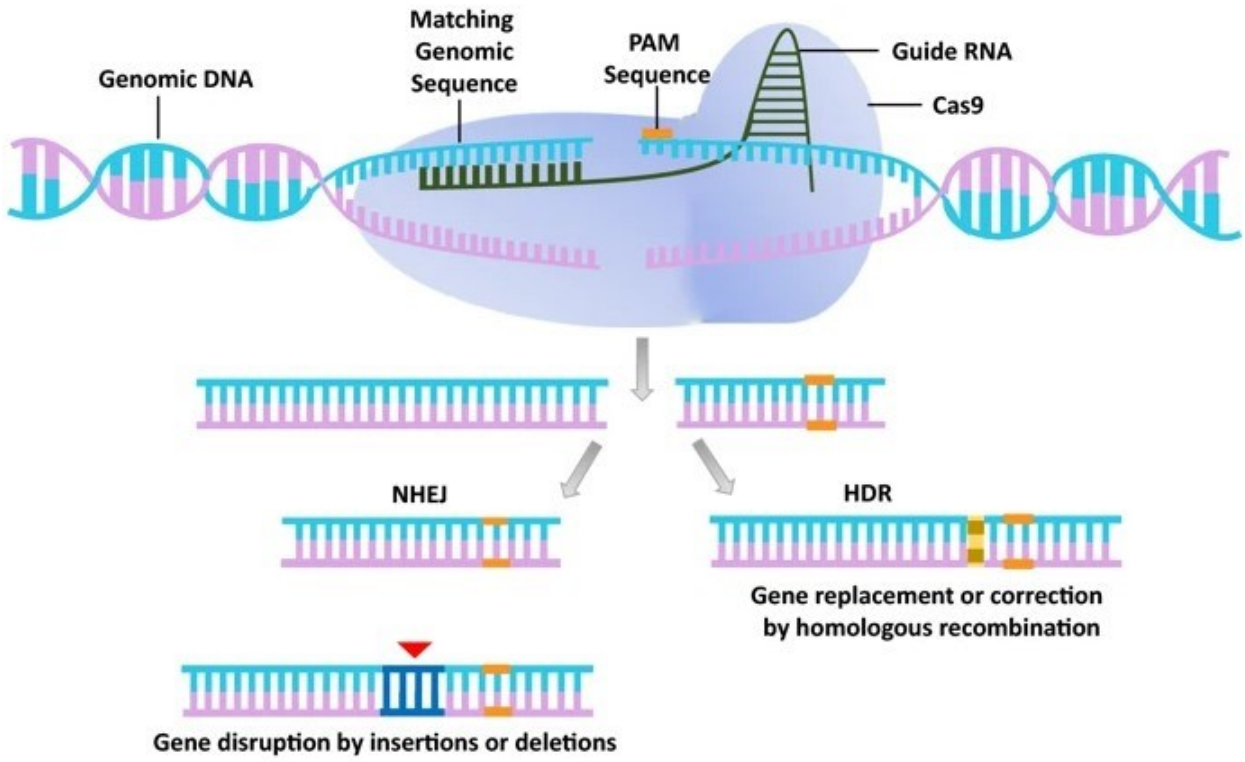



**Figure 3. Schematic of CRISPR–Cas9-mediated genome editing. The *Streptococcus pyogenes* derived CRISPR–Cas9 RNA-programmable DNA endonuclease is targeted to a DNA sequence via a single guide RNA (sgRNA) sequence, which base-pairs with a 20-nt DNA sequence upstream of the protospacer-associated motif (PAM), resulting in a 3-bp double-strand break (DSB) upstream of the NGG. Random indels or precise modifications introduced into the genomic DNA by the NHEJ or HDR pathway. (reproducted from Gosh et al., 2019)**

However, due to the complex nature of abiotic stress, fewer genome-editing studies have been conducted in this research area. The identification of regulatory genes involved in the abiotic stress tolerance will provide new molecular targets to implement breeding programs. In particular, transcription factors (TFs) represent an important class of targets for modulating the level of expression of several downstream genes and activating many stress signals (Zafar et al., 2019). Various TF members belonging to the MYB, WRKY, NAC, DOF, DREB, AREB/ABF (ABA response-element binding factor), GBFs (G-box binding factors), and AP2/ERF families (Bostock et al., 2014) have been characterized for their involvement in abiotic stress. Recently, in wheat protoplasts, two abiotic-stress-responsive transcription factor genes encoding dehydration responsive element binding protein 2 (TaDREB2) and ethylene responsive factor 3 (TaERF3) were edited conferring adaptive abiotic stress responses (Kim et al., 2018). Rice transcription factor OsNAC041 have also been targeted to increase salinity tolerance (Bo et al., 2019). These findings confirmed that CRISPR/Cas9 could be the future of targeting regulatory genes of complex quantitative traits related to abiotic stress.

## 1.4 SOLANUM LYCOPERSICUM AS A MODEL FOR HEAT STRESS

Tomato (*Solanum lycopersicum*) is the most important vegetable and the seventh most valuable crop worldwide. It possesses unique properties, offering both a rich source of vitamin A (lycopene) and C (ascorbate), and antioxidants that have been shown to correlate with prevention of cancer and cardiovascular diseases (Rao and Agarwal, 2000). Tomatoes are native to South America and were introduced in Europe in the 16th century. On global scale, the annual production of fresh tomatoes amounts to approximately 160 million tonnes (Mt) (http://faostat3.fao.org/home/index.html). The biggest tomato producers are found in Asia, which represents 60.3% of tomato production. Europe represents 12.7% of the world tomato production (figure 4). Recent data reported by AMITOM (Mediterranean International Association of the Processing Tomato) show that Italy is the top European producer with 4.750 Mt (figure 5).

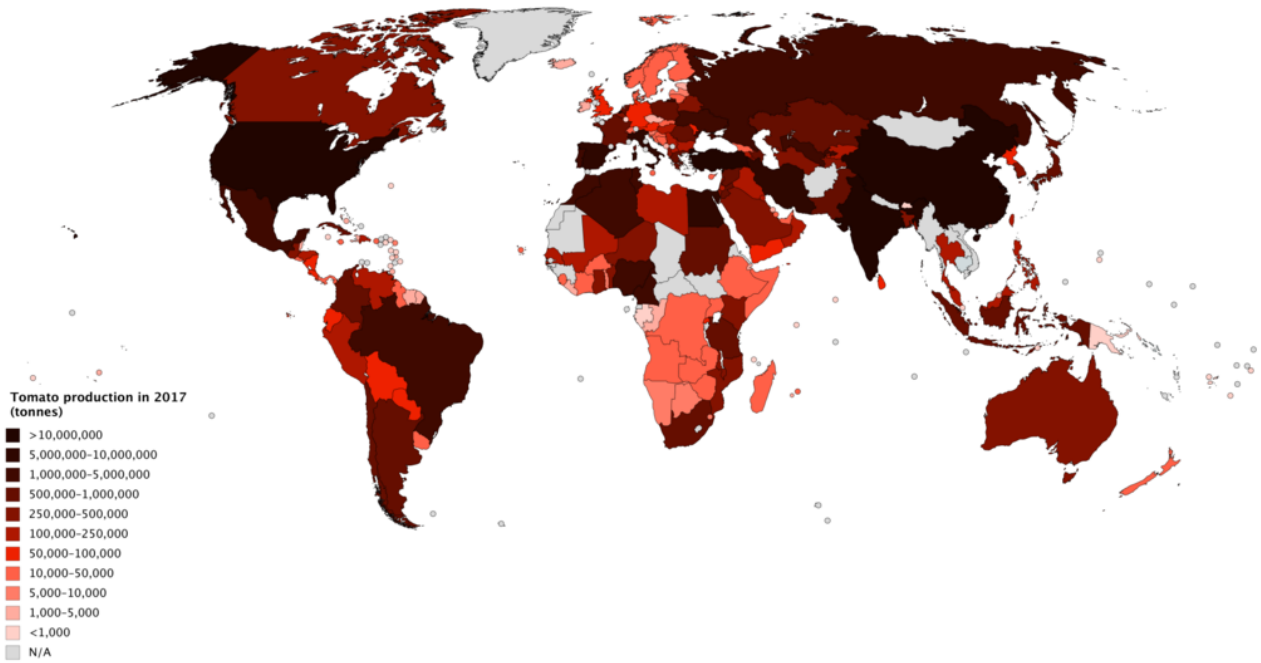



**Figure 4. Choropleth map showing countries by tomato production in tonnes, based on data from the Food and Agriculture Organization Corporate Statistical Database (FAOSTAT, http://faostat3.fao.org/home/index.html).**

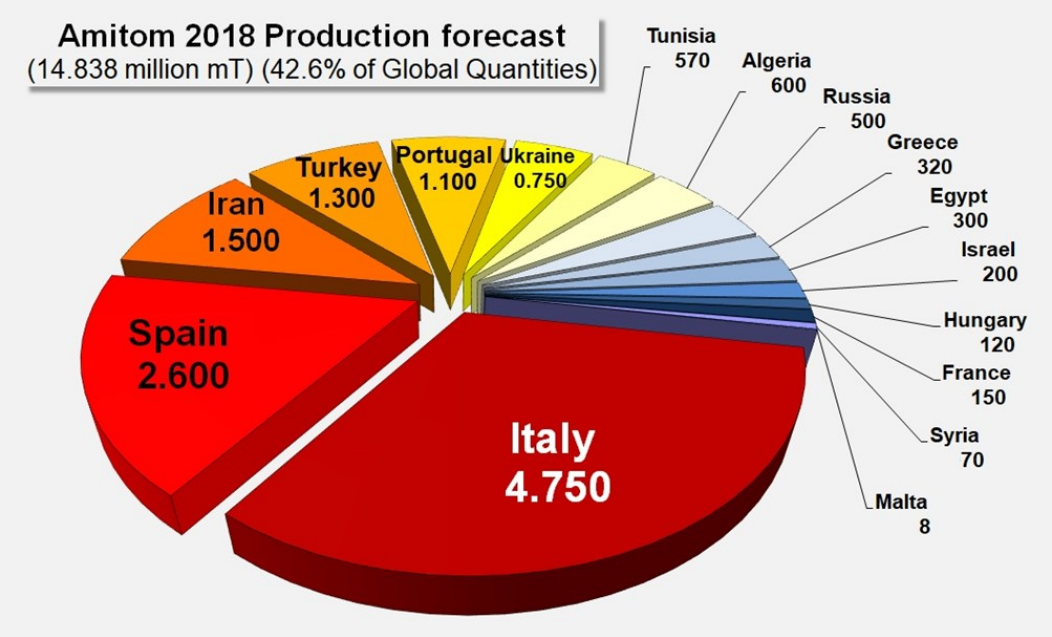



**Figure 5. AMITOM (Mediterranean International Association of the Processing Tomato) 2018 tomato production.**

Besides to have invaded the crop market and our plates, tomato is considered by researchers as a model crop due to its attractive traits (Foolad, 2007): a relative short life cycle, a high self-fertility that allows genetic stability over generations, a recently sequenced genome and the availability of several genomic resources.

The optimal growing temperature of tomato is between 21 ◦C to 30 ◦C during the day and 18.5 ◦C to 21 ◦C during the night (Jones, 2008), and cultivation of tomatoes under higher temperatures affects plant growth and yield (Sato et al., 2000; Gerszberg et al., 2017). Indeed, a temperature of a few degrees above this threshold can lead to serious deleterious effects, such as flower abscission, decrease of pollen quality, abnormal growth, reduced fruit set and a huge loss in yield (Hasanuzzaman et al., 2013). Hence, the increased temperatures forecasted for the near future (e.g. increase of heat waves), the importance of the tomato crop worldwide, the variation in heat stress response of different tomato accessions and the applicability of tomato as a model plant, makes it a valuable object for the study of the mechanisms of heat tolerance of the whole plant and, in particular, of the male reproductive organs, whose high sensitivity is the bottleneck of fruit production under high temperatures.

## 1.5 OBJECTIVIES OF THE THESIS

The main goal of this thesis was the exploitation of genetic and genomic resources for accelerating the development of tomato cultivars with improved yield under high temperature conditions.

In Chapter II, a tomato F4 segregating population was phenotypically evaluated for quantitative and qualitative traits under heat stress condition to identify superior genotypes. To better investigate the basis of heat stress tolerance selected individuals were also assessed for pollen viability. Finally, the cultivar Moneymaker was tested for a local heat treatment to understand if the heat stress response is a local or a systemic process.

In Chapter III, Genomic prediction models for yield production per plant (YP) and soluble solid content (SSC) based on genotyping by sequencing (GBS) data were set up and tested on segregating populations in order to increase the breeding gain. A QTL analysis to better understand the genetic architecture of heat tolerant traits in tomato was also performed.

In chapter IV, the SlDof transcription factors family was investigated to identify potential regulators of temperature stress tolerance in tomato. A phylogenetic analysis of this gene family including Arabidopsis members was performed to identify a suitable candidate for a CRISPR/Cas9 experiment.

# CHAPTER II

# TOMATO PLANTS PHENOTYPING UNDER HEAT STRESS CONDITIONS

## 2.1 INTRODUCTION

The elucidation of plant response to abiotic stresses has been an important goal for plant scientists and breeders since many decades. Indeed, crop yield reduction due to climatic events is threatening global food security (Bailey-Serres et al., 2012; Tele Ayenan et al., 2019). Among abiotic stresses, high temperature is one of the major affecting plant reproduction and productivity (Dane et al., 1991). Tomato *(Solanum lycopersicum)* is one of the most important worldwide horticulture crop despite being sensitive to high-temperature. The optimum temperature range for tomato is of 21 ◦C to 30 ◦C during the day and 18.5 ◦C to 21 ◦C during the night (Jones B.J, 2008), and cultivation of tomatoes under higher temperatures affects plant growth and yield (Sato et al., 2000; Gerszberg and Hnatuszko-Konka K, 2017).

Developing heat stress tolerant tomato cultivars may be a valuable strategy to cope with climate changes. To date, screening for thermo-tolerance is undertaken in various environments such as growth chambers, greenhouse, and in open field (Mesihovic et al., 2016). Several studies have focused on open field screening for the identification of tomato cultivars tolerant to high temperature in order to reproduce the real crop conditions (Dane et al 1991; Kartikeya et al., 2012; Kugblenu et al., 2013; Zhou et al., 2015; Singh et al., 2015). However, several uncontrollable factors (biotic stress, variation in soil type, fertility level, etc.) that can affect plant response should be taken in account (Mesihovic et al., 2016).

Different quantitative and qualitative parameters such as pollen viability, fruit set, yield productivity per plant, total number of fruits, fruit shape and SSC (soluble solid content) of fruits, may be assessed to evaluate the high temperature plant tolerance (Wahid et al., 2007). Among these, fruit set and yield have been the main targeted traits for most heat tolerance screening in tomatoes (Kugblenu et al., 2013; Singh et al., 2015) and has been shown that the decrease in fruit set and yield in tomato under long-term mildly elevated temperatures is correlated with a decrease in pollen viability (Sato et al., 2000; Pressman et al., 2006; Xu et al., 2017). In the last years, the research on heat tolerance mechanisms was focused on heat tolerance sub-traits such as pollen number per flower, pollen viability, number of flowers per plant, number of inflorescences per plant (Yan et al., 2010; Weerakoon et al., 2008; Driedonks et al., 2018). The indirect assessment of fruit setting

and/or yield through these sub-traits is mainly supported by the complexity of these traits and the relative simplicity of the sub-traits genetic architecture (Paupière et al., 2017; Driedonks et al., 2018).

At the light of these observations and to better investigate the basis of heat stress tolerance, a segregating population obtained from the tomato variety JAG8810 was characterized at phenotypic level for nine traits under heat stress in field. This variety is very interesting because it was previously selected for its high tolerance under heat stress by Bayer seed company. The agronomic evaluation of JAGF4 population allowed us to identify 8 extreme individuals, which have been also phenotyped in growth chamber (with full control of temperature, light and humidity) for pollen viability, flowers and inflorescences number per plant. In addition, a comparison between local and global heat treatment on the cultivar *Moneymaker* in order to improve our knowledge about heat stress effects on plant tissues and organs, was carried out. Furthermore, the Tomato Analyzer tool (Rodrıguez et al., 2011) was tested for the analysis of fruit shape of all 100 JAGF4 lines to complement the previous agronomic characterization with digital fruit descriptors.

## 2.2 MATERIALS AND METHODS

### 2.2.1 Material and plant grown

The tomato variety JAG8810, kindly obtained from Bayer station (Latina, Italy) was self-pollinated from F2 to F4 generations through a Single Seed Descendent (SSD) schema in greenhouse, of Department of Agriculture of the University of Naples Federico II (N 40° 48.' 49.352''; E 14° 20' 40.073''). In 2017, 100 F4 lines were grown in open field in Southern Italy region (Battipaglia-Campania; N 40° 58.' 56.69''; E 14° 96' 10.02''), usually characterized by high temperatures during the flowering and fruit set periods (from June to August). Fifteen seeds for line were sown in plateau under plastic-house. Ten seedlings for single JAGF4 line in a completely randomized design were transplanted in field under plastic tunnel in the first half of May in order to expose plants to high temperatures. Tomato plants were grown following the standard cultural practices of the area and temperature and climatic data were recorded using the weather station Davis weatherlink.

### 2.2.2 Trait evaluation and statistical analysis

In the experimental field, three random plants per genotype were analyzed for traits related to flowering, fruit production and quality. For four of these traits (fruit earliness-FRL, leaf coverage-LC, inflorescences number-IN and contemporaneous ripening-CR) a Selection Index (SI) was calculated by assigning an arbitrary scale scored on a 1 to 5. Furthermore, since many genotypes have suffered a viral attack during their grown, all genotypes were also scored for disease symptoms presence (0 for absence, 1 for presence). The percentage of fruit set (FS) were evaluated on inflorescences produced from the second to the fifth truss on three plants per line randomly chosen. Finally, at fruit red ripe stage, total fruit number (TFN), soluble solid content (SSC-3 fruits per plant) and yield production per plant (YP) were measured. Correlations between phenotypic traits were estimated using the Spearman coefficient analysis. Principal component analysis (PCA) combining all phenotypic traits was conducted to determine the origin and structure of variation as well as contribution of the observed characteristics in total variability.

### 2.2.3 Fruit characterization by Tomato Analyzer software and statistical analysis

Three ripe fruits per genotypes were longitudinally cut and scanned with a photo scanner (Epson, V39) at a resolution of 300 dpi and subjected to morphometric analysis with Tomato Analyzer version 3 software (Rodríguez et al., 2010; Strecker et al., 2010). Data were recorded for a total of 9 categories of Tomato Analyzer descriptors, which included basic (7), fruit shape (3), homogeneity (3), blockiness (3), distal fruit end shape (4), proximal fruit end shape (4), internal eccentricity (5), asymmetry (6), and latitudinal section (3) descriptors. Default settings were used for blockiness and proximal fruit end shape and distal fruit end shape descriptors (Rodríguez et al., 2010). Data analyses of all descriptors were performed using standard parametric statistics (Little and Hills, 1978). Mean and range values were calculated for each descriptor.

### 2.2.4 Pollen viability assessment

All experiments described in this section were conducted to assess the pollen viability of selected tomato lines at the Molecular Plant Physiology Department- Radboud University- Nijmegen (Netherlands). A preliminary experiment (BACTH 1) was performed to set up the optimal temperature range using the variety JAG8810 (December 2018). A second experiment (January- April 2019) was conducted with JAG8810 and eight F4 extreme individuals deriving from JAG8810 (BACTH 2). Seeds from both BATCHs were incubated in 2.5% hypochlorite for 30 min at room temperature before sowing them in soil (Horticoop, Lentse Potgrond, Slingerland Potgrond) covered with vermiculite (Agra - Vermiculite) under standard greenhouse condition (25˚C/19˚C day/night) and humidity (70/80% day/night). Seedlings were transferred to 0.5 L pots after two weeks and, after one month, placed in 12 L pots. When the transition from the vegetative to the generative phase occurred, flower buds were removed, and the plants were transferred to a climate chambers using a randomized design. In particular, BATCH 1 plants were transferred in 3 different chambers (5 plants per chamber) maintaining a 14/10 h day/night photoperiod and humidity of 70/80% at three control temperatures: 30.5˚C/25˚C, 32˚C/26˚C, 33.5˚C/27˚C day/night. Whereas, 3 different experiments with BATCH 2 were conducted using the same photoperiod and humidity conditions of BATCH 1 and a control temperature of 33˚C/27˚C. In each BACTH 2 experiment, 3 plant per genotypes were used.

The pollen viability (PV) screening analysis was conducted on newly opened flowers, collected each morning for four days from 9 to 11 am under the heat stress. In total, three flowers per plant were collected and three plants were used per genotype. Flowers were collected into a petri dish filled with a wet paper. From each flower, anthers were cut into three to four pieces with a razor blade on a glass and transferred in a 1,5 ml Eppendorf. One microliter of AmphaFluid 4 Buffer was added, vortexing

for 5 seconds. Measurements were carried out, loading 500µl of the solution through a microfluidic chip, that permits a single cell analysis by impedance flow cytometry AmphaTMZ30 (Amphasys). This analysis allows to detect dead and viable pollen exploiting measurements in a broad radiofrequency (RF) range (0.1 – 30 MHz) and to determine size, membrane capacitance and cytoplasmic conductivity of cells. In BATCH 1, each replicate was represented by a pool of pollen derived from 3 flowers collected from the same plant. For each experiment the average and the standard deviation of 5 biological replicates was calculated. In BATCH 2, three biological replicates were used for each single experiment. Furthermore, the total number of inflorescences (TNI) and flowers (TNF) were also counted to correlate the PV to TNI and TNF.

### 2.2.5 Setting up a method to evaluate heat stress effects

To test if the effect of high temperatures in plant is local or systemic, an experiment, in which an isolated plant branch was treated with a different temperature from the rest of the plant, was set up. The cultivar *Moneymaker* was used for the experiment because it shows high sensitiveness to the heat. In order to isolate a single branch of the plant from the rest of the room, a hole in the wall of a climate room connecting the outside corridor to the grown chamber, has been exploited. Different trials were carried out to set up the most suitable conditions. At first, a tube connecting an air pump located in the corridor (at the temperature of 26°C) to the plants' branch, isolated through a plastic bag from the rest of climate room at 33 °C, was tested. However, the tubing system didn't allow to maintain the temperature stable. Therefore, a refrigerator at 4.6 °C to cool the tube before connecting it to the isolated branch was added (figure 1). Before starting the biological experiment, the temperature in the plastic bag was measured every day for one week through a thermometer. In addition to the external pump, another air pump, located into the climate room at 33°C degree, was also connected to a plastic bag containing another plant branch of the same plant as positive control. After 2 weeks of stress the PV was measured as described in the section 2.2.4. In particular, for four days one flower from the bag connected with the corridor (at 26°C), one flower from the bag connected to the air pump into the room at 33°C, one flower from the same plant in the grow chamber at 33°C (without bag) and 3 flowers deriving from 3 different Moneymaker plants located in another climate room at 25°C (control), were analyzed.

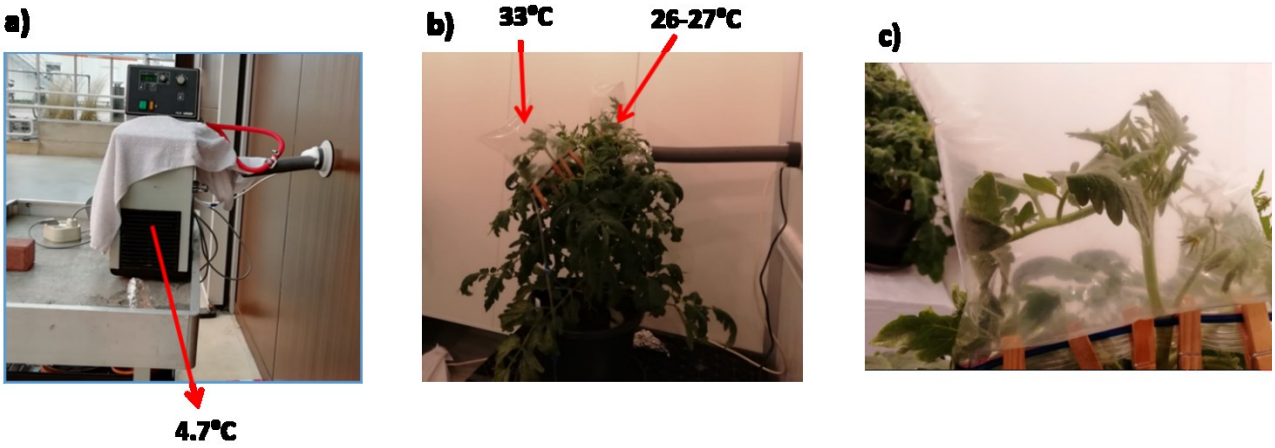



**Figure 1. Setting up the experiment.**

### 2.2.6 Statistical analysis

Correlation analysis (Spearman or Pearson) and PCA analysis were processed in the statistical software SPSS statistics 24.0 (IBM Corp., Armonk, NY, USA).

All statistical analyses for pollen viability were performed using a logit transformation data value'=LN((value+1)/(101-value)). Analysis of significant differences among replicates and experiments was performed running a T-test in Excel. The correlation coefficients among phenotypic traits were calculated by a Pearson correlation analysis for PV, TNF, TNI and YP using SPSS software.

## 2.3 RESULTS

### 2.3.1 Analysis of phenotypic traits in tomato segregating population for heat stress tolerance

In this study, 100 F4 tomato genotypes obtained from the heat stress tolerant tomato variety JAG8810, were grown under high temperatures. Tomato plants grown in the South of Italy are usually transplanted in open fields in April. However, in this study the genotypes were transplanted in open fields under plastic tunnel with a delay of one month compared to the usual transplanting period, thus imposing a high-temperature condition during the reproductive stages. Figure 2 reports the variation of maximum temperatures from May until the end of July. Temperatures mostly resulted upper than 30°C and therefore we considered them sufficiently elevated to analyze the response to high temperature conditions of the different genotypes. Indeed, maximum temperatures of 35°C during the day, represent an extremely critical threshold in the sensitive stages of reproductive development (Jones B.J, 2008). In our trials, former temperature, was frequently exceeded (11 days) (figure 2).

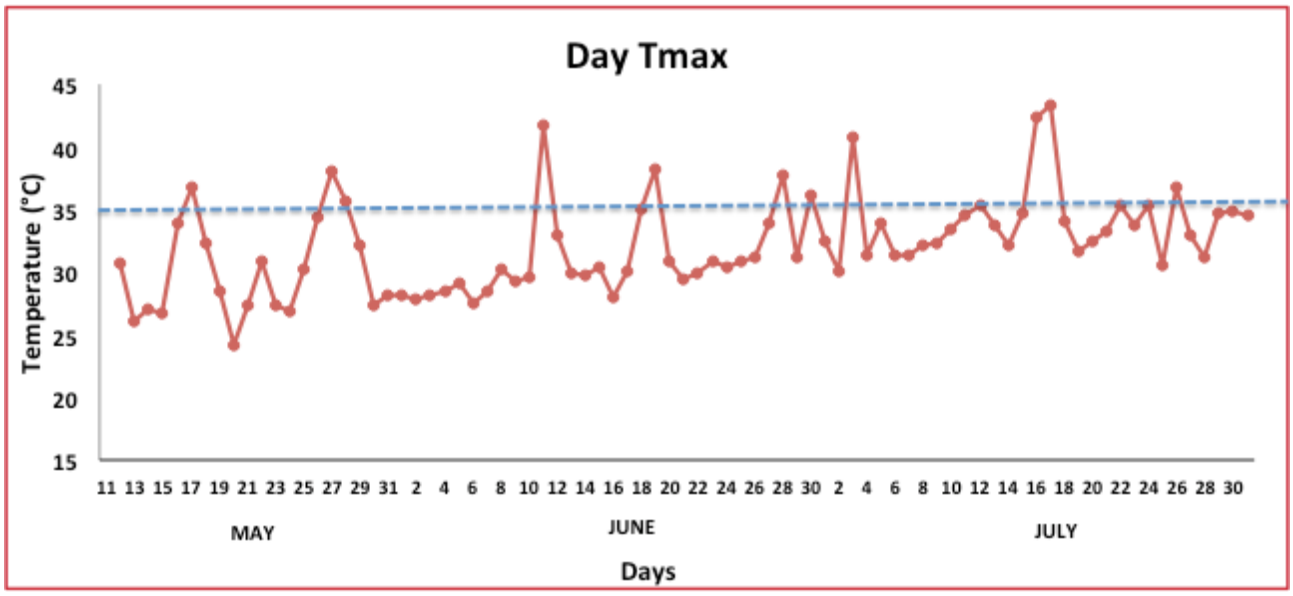



**Figure 2. Maximum temperatures recorded in the experimental fields (Battipaglia, It) during the day from May to July 2017.**

In these extreme conditions 100 F4 genotypes were phenotyped for 9 traits related to flowering, fruit set, soluble solid content and productivity (supplementary table 1). Table 1 reports the average and

the minimum and maximum value of 4 measured traits (FS, TFN, YP and SSC). As we can see, we found high variability for analyzed traits related to flowering and fruit production. For example, YP ranged between a minimum value of 0.8 kg to a maximum value of 16 kg; FS (%) ranged between a minimum of 8 to a maximum of 78; SSC ranged between a minimum of 3 to a maximum of 6.8.

| Trait | Mean | Min | Max |
|---|---|---|---|
| FS (%) | 38.3 | 8 | 78 |
| TFN | 156.7 | 32 | 354 |
| YP (Kg/plant) | 7.9 | 0.8 | 16 |
| SSC | 4.4 | 3 | 6.8 |

**Table 1. Mean value, minimum value and maximum value of traits related to flowering, yield and SSC. FS (%): fruit set percentage; TFN: total fruit number per plant; YP: yield production per plant; SSC: Soluble solid content.**

A correlation matrix was obtained using all available data (table 2). A positive and significant correlation among parameters relative to plant production was found. Indeed, YP was positively and strongly correlated to TFN (r= 0.888; P<.0.01) and FS (r=0.576; P<0.01). The yield contributing trait such as YP was also positive significantly correlated to IN (r=0.373; P<0.01) and FRL (r=0.326; P<0.05) whereas, it was negative correlated with CR (r=-0.531; P<0.01) and SSC (r=-0.314; P<0.01). SSC was also negative correlated with TFN (r=-0.267; P<0.05), IN (r=-0.306; P<0.01) and FRL (r=-0.340; P<0.01). However, we found a slight positive correlation between CR and SSC (r=0.273; P<0.05).

| | YP (Kg) | TFN | SSC | CR | FS | IN |
|---|---|---|---|---|---|---|
| **TFN** | .888** | | | | | |
| **SSC** | -.314** | -.267* | | | | |
| **CR** | -.531** | -.451** | .273* | | | |
| **FS** | .576** | .598** | | -.234* | | |
| **IN** | .373** | .343** | -.306** | -.24* | .278* | |
| **LC** | | | | -.515** | | |
| **FRL** | .326* | .236* | -.34** | | | .278* |

**Table 2. Correlation coefficient (r) among variables. *P<0.05, **P<0.01. YP: yield production per plant; TFN: total fruit number per plant; SSC: Soluble solid content; CR: contemporaneous ripening; FS: percentage of fruit set; IN: inflorescence number; LC: leaf coverage; FRL: fruit earliness.**

Finally, in order to select the best and worst performing genotypes under high temperatures a PCA analysis based on all phenotypic analyzed data was carried out. The first three PCA components accounted for 69.033% of total variation among genotypes means (table 3). In particular, the first PCA component explained 37.5% of the total variation among the genotypes and the most important traits for PCA1 were YP, TFN and FS (table 4). PCA2 explained 18.76 % of the total variability among genotypes. The most important traits for PCA2 were CR, viroses and LC (table 4). Finally, PCA3 explained 12.77 % of the total variation. SSC, IN and FRL represents the most important traits for PCA3 (table 4). The PCA scores of genotypes were plotted with respect to PCA1, PCA2 and PCA3 to assess the genotypic differences. Great variations were found among tomato lines. This analysis allowed us to identify the best and worst performers under heat stress for traits related to plant production (figure 3).

| | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| Component | Total | Variance % | Cumulative % | Total | Variance % | Cumulative % | Total | Variance % | Cumulative % |
| 1 | 3.375 | 37.500 | 37.500 | 3.375 | 37.500 | 37.500 | 2.383 | 26.480 | 26.480 |
| 2 | 1.689 | 18.763 | 56.263 | 1.689 | 18.763 | 56.263 | 2.236 | 24.843 | 51.323 |
| 3 | 1.149 | 12.770 | 69.033 | 1.149 | 12.770 | 69.033 | 1.594 | 17.710 | 69.033 |
| 4 | .797 | 8.855 | 77.887 | | | | | | |
| 5 | .705 | 7.829 | 85.717 | | | | | | |
| 6 | .486 | 5.404 | 91.121 | | | | | | |
| 7 | .378 | 4.204 | 95.325 | | | | | | |
| 8 | .337 | 3.747 | 99.072 | | | | | | |
| 9 | .083 | .928 | 100.000 | | | | | | |

**Table 3. Total variance explained in PCA analysis**

| | Component | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| YP | **.843** | -.366 | .261 |
| TFN | **.864** | -.297 | .141 |
| SSC | .022 | .128 | **-.794** |
| CR | -.239 | **.783** | -.177 |
| Virosis | -.136 | **.745** | -.299 |
| FS | **.841** | .129 | .043 |
| IN | .170 | -.174 | **.680** |
| LC | -.050 | **-.857** | -.184 |
| FRL | .332 | .219 | **.506** |

**Table 4. PCA characteristics of 100 tomato genotypes analyzed. Note: Bold values represent parameters important for PCA1, PCA2, PCA3. YP: yield production per plant; TFN: total fruit number per plant; SSC: Soluble solid content; CR: contemporaneous ripening; Virosis; FS: percentage of fruit set; IN: inflorescence number; LC: leaf coverage; FRL: fruit earliness.**

**PCA 1**: YP, TFN, FS%
**PCA 2**: CR, viroses, LC
**PCA 3**: SSC, IN, FRL

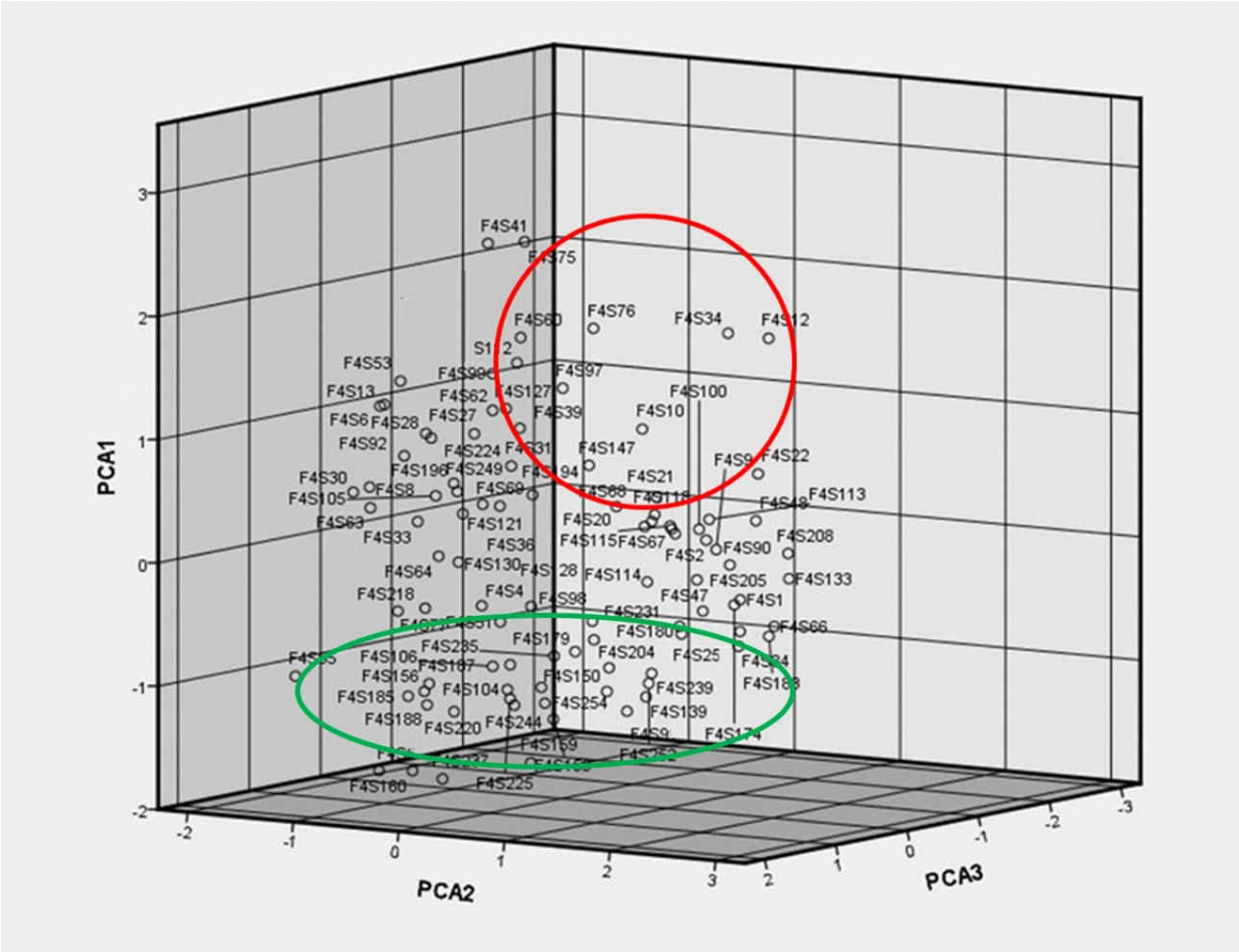



**Figure 3. Three-dimensional PCA plot of 100 genotypes analyzed for 9 traits (YP: yield production per plant; TFN: total fruit number per plant; SSC: Soluble solid content; CR: contemporaneous ripening; FS: percentage of fruit set; IN: inflorescence number; LC: leaf coverage; FRL: fruit earliness). The best genotypes are highlighted in the upper part of graphic by a red circle; the worst are highlighted by a green circle.**

### 2.3.2 Fruit characterization by Tomato Analyzer

An electronic scanning of fruits was performed to digital phenotype to accurately analyze the tomato fruit of JAGF4 population (table 5). For 38 tomato fruit descriptors, unit, mean and range were reported. Most of traits studied were dissected in sub descriptor categories that showed wide variation. In this respect, the descriptors with largest variation were the Area (1487.8-3455.2 mm2) within the *Basic descriptors* category; Curved Fruit Shape Index (0.79-1.74) within the *Fruit shape index descriptors*; Fruit Shape Triangle (0.63-1.37) within *Blockiness descriptors*; Circular (0.02-0.14) in *Homogeneity descriptors* category, Proximal Angle Micro (58.8-187.9) within *Proximal fruit end shape descriptors;* Distal Angle Micro (54.5-325.7) in *Distal fruit end shape descriptors* category; H. Asymmetry.ob (0.00–0.3) in *Asymmetry descriptors* category; Fruit Shape Index Internal (0.7–1.56) within *Internal eccentricity descriptors*, and Lobedness Degree (1.7-9.9) in *Latitudinal section descriptors* category (table 5; supplementary figure 1). For eight descriptors of the categories *Distal fruit end shape* (Distal indentation area and distal end protrusion), *Asymmetry* (Obovoid, Ovoid, H.Asimmetry.Ov and H.Asymmetry.Ob) and *Proximal fruit end shape* (Shoulder height and Proximal Indentation Area) values of 0.00 were observed (table 5).

| Descriptors | Units | Mean | Range |
|---|---|---|---|
| *Basic descriptors* | | | |
| Perimeter | Mm | 183.57 | 151.6-235.73 |
| Area | mm2 | 2178.17 | 1487.81-3435.19 |
| Width Mid-height (WMh) | Mm | 53.15 | 39.2-67.42 |
| Maximum Width (MW) | Mm | 54.10 | 40-68.8 |
| Height Mid-width (HMw) | Mm | 50.14 | 39.32-72.16 |
| Maximum Height (MH) | Mm | 50.95 | 39.33-74.17 |
| Curved Height (CH) | Mm | 53.90 | 41.75-74.5 |
| *Fruit shape index descriptors* | | | |
| Fruit Shape Index External I (FSIE I) | | 0.96 | 0.68-1.52 |
| Fruit Shape Index External II (FSIE 2) | | 0.96 | 0.69-1.56 |
| Curved Fruit Shape Index (CFSI) | | 1.03 | 0.79-1.74 |
| *Blockiness descriptors* | | | |
| Proximal Fruit Blockiness (PFB) | | 0.66 | 0.45-0.79 |
| Distal Fruit Blockiness (DFB) | | 0.67 | 0.51-0.77 |
| Fruit Shape Triangle (FST) | | 0.98 | 0.63-1.37 |
| *Homogeneity descriptors* | | | |
| Ellipsoid | | 0.03 | 0.01-0.06 |
| Circular | | 0.07 | 0.02-0.14 |
| Rectangular | | 0.53 | 0.48-0.58 |
| *Proximal fruit end shape descriptors* | | | |
| Shoulder Height (SH) | | 0.015 | 0.00-0.03 |
| Proximal Angle Micro (PAMicro) | Degree | 147.94 | 58.8-187.9 |
| Proximal Angle Macro (PAMacro) | Degree | 150.00 | 78.21-173.56 |
| Proximal Indentation Area (PIA) | | 0.03 | 0.00-0.03 |
| *Distal fruit end shape descriptors* | | | |
| Distal Angle Micro (DAMicro) | | 155.10 | 54.49-325.74 |
| Distal Angle Macro (DAMacro) | Degree | 152.47 | 75.91-200.88 |
| Distal Indentation Area (DIA) | Degree | 0.01 | 0.00-0.09 |
| Distal End Protrusion (DEP) | | 0.01 | 0.00-0.19 |
| *Asymmetry descriptors* | | | |
| Obovoid | | 0.075 | 0.00-0.22 |
| Ovoid | | 0.057 | 0.00-0.23 |
| V. Asymmetry (V.A) | | 0.097 | 0.02-0.26 |
| H. Asymmetry.ob (H.A.ob) | | 0.054 | 0.00-0.3 |
| H. Asymmetry.ov (H.A.ov) | | 0.033 | 0.00-0.16 |
| Width Widest Pos (WWP) | | 0.488 | 0.34-0.62 |
| *Internal eccentricity descriptors* | | | |
| Eccentricity | | 0.79 | 0.74-0.8 |
| Proximal Eccentricity (PE) | | 0.89 | 0.88-0.89 |
| Distal Eccentricity (DE) | | 0.89 | 0.86-0.89 |
| Fruit Shape Index Internal (FSII) | | 0.96 | 0.69-1.56 |
| Eccentricity Area Index (EAI) | | 0.39 | 0.33-0.43 |
| *Latitudinal section descriptors* | | | |
| Lobedness Degree (LD) | | 3.70 | 1.73-9.86 |
| Pericarp Area (PA) | | 0.57 | 0.56-0.62 |
| Pericarp Thickness (PT) | | 0.25 | 0.24-0.25 |

**Table 5. Tomato Analyzer descriptors data. Unit used, mean and range variation observed in the 100 tomato JAGF4 lines are reported.**

Finally, with the aim to find correlations among descriptors, a correlation matrix was created (figure 4). Results highlight a strongly positive and significant correlation for 5 (out of 7) descriptors related to fruit size (basic descriptors). A strongly positive correlation was also found for the three fruit shape index descriptors that describe the external form of the fruit. Instead negative correlations were found among the three Fruit shape index descriptors and 2 basic descriptors (Width mid-height and

Maximum Width). Positive correlation has been found between the three fruit shape index descriptors and internal fruit shape index belonging to the category of Internal eccentricity descriptors. Interestingly, both fruit shape index descriptors and internal eccentricity descriptors resulted negatively correlated to the 2 basic descriptors related to fruit size (Width mid-height and Maximum Width). Other important positive correlations were found among Lobediness degree descriptor (Latitudinal section) and Circular descriptor (Homogeneity) and between 2 descriptors belonging to the same class, Ovoid and H. Asymmetry. Ob (Asymmetry descriptors). Moreover, a strong negative correlation was found between H. Asymmetry. Ob (Asymmetry descriptors) and Fruit Shape triangle descriptor belonging to the class Blockiness descriptors. Supplementary figure 2 reports the significant correlation values (P<0.05) found.

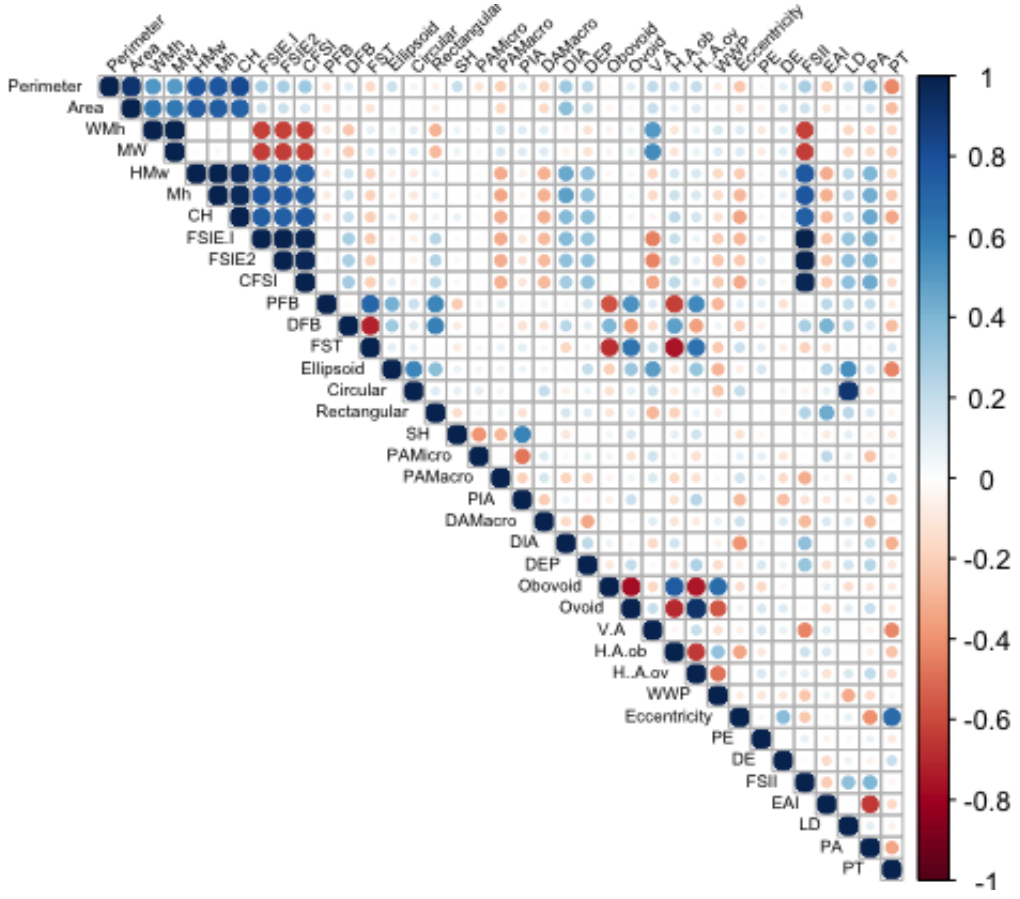



**Figure 4. Pearson's rank correlation coefficients between pairs of descriptors. Only correlation coefficients with P value<0.01 after Bonferroni correction are shown. Color intensity is proportional to the correlation coefficients. On the right side of the correlogram, the legend color shows the correlation coefficients and the corresponding colors.**

### 2.3.3 Characterization of pollen traits in extreme phenotypes for heat stress tolerance

As described above (section 2.3.1), the PCA analysis based on the combination of 9 phenotypic traits analyzed in the field during the summer 2017 allowed us to select the best and worst performers in the investigated population. Among these, 8 extreme JAGF4 genotypes (4 tolerant and 4 susceptible for traits related to production) were selected and phenotyped for pollen viability through a flow cytometer (table 6).

| | Line | YP | TFN | FS |
|---|---|---|---|---|
| Worst | S132 | 4 | 70 | 15 |
| | S185 | 6.5 | 77 | 24 |
| | S153 | 2.3 | 45 | 27 |
| | S239 | 4 | 80 | 20 |
| Best | S75 | 13.5 | 330 | 73 |
| | S76 | 13.7 | 333 | 53 |
| | S196 | 12.5 | 249 | 50 |
| | S99 | 12 | 208 | 65 |

**Table 6. Summary of agronomic characteristics of 8 JAGF4 extreme genotypes. Worst represents the JAGF4 susceptible genotypes. Best represents JAGF4 tolerant genotypes. Value of YP, TFN and FS: were reported.**

To set up the optimal temperature condition for the PV test of lines, a preliminary experiment (BATCH 1) using the hybrid JAG8810 was conducted. The percentage of PV varied from 72.15 at 30.5˚C, 68.9 to 22.8 at 33.5˚C (figure 5). Results showed that there were no statistically significant differences between the % PV at 30.5˚C and 32˚C, while the plant at 33.5˚C ($P<0.001$) showed a PV reduction of 50 % compared to the plants at 30˚C and of 46% compared to the plants at 32˚C ($P<0.001$).

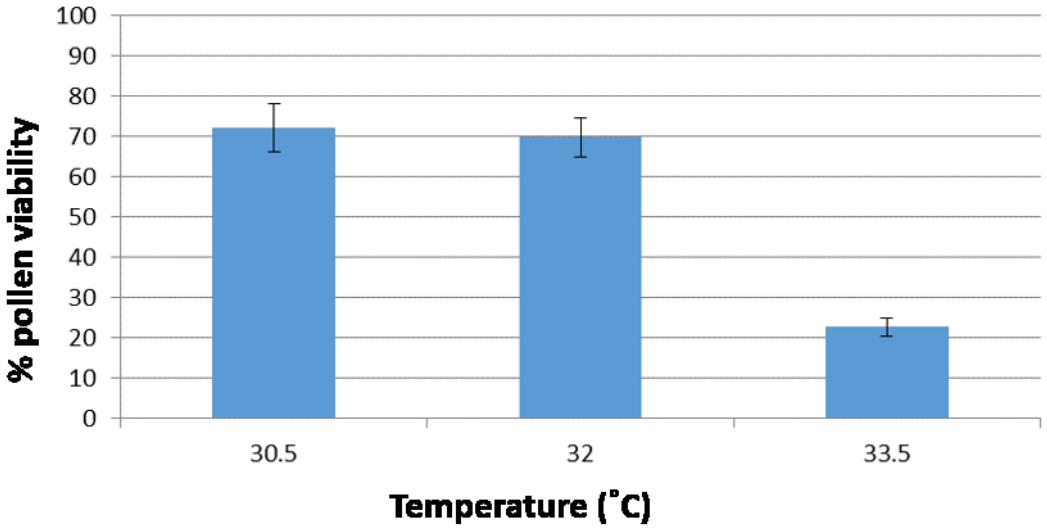



**Figure 5. Comparison of PV in JAG8810 at three different temperatures. Values represent the mean ± standard deviation.**

At the light of these results we decided to use a heat stress temperature of 33˚C/27˚C (day/night) for the further PV screening (BATCH 2). The percentage of PV under high temperature showed high variability among genotypes, also compared to the control (JAG8810). All differences between each JAGF4 genotype and JAG8810 resulted statistically significant (figure 6a). All worst lines had lower PV compared to the control whilst S76, a good performer line, showed a PV almost 10 times higher than the control (figure 6a).

Based on genotypes response to the three parameters tested (PV and TNI, TNF), 3 classes, using the control as reference were identified: (i) genotypes with higher PV and higher TNI and TNF (S76), (ii) genotypes with lower PV but higher TNI and TF (S75, S99, S196 and S132), genotypes with lower PV and lower TNI and TNF (S185, S153, S239) (figure 6a;6c). These data allowed us to identify the genotype S76 as one of the best performer in the studied population. It performed also better than JAG8810 both for traits related to productivity in field (section 2.3.1) and for PV, TNI and TNF. Furthermore, our results concerning the total number of inflorescences (TNI) and flowers (TNF) counting, showed that genotypes with high TNI had also higher TNF since they were strongly correlated ($r=0.961$; $P<0.01$). Our results showed also a positive and significant correlation between TNF and PV ($r=0.443$; $P<0.01$) but not between TNI and PV (table 7). Moreover, with the aim to verify if also the yield contributing trait such as YP were correlated to TNF, TNI and PV, a correlation analysis including values of YP for the 8 genotypes analyzed was performed. Interestingly, the correlation matrix highlight that YP was positively and highly correlated to PV ($r= 0.642$; $P<0.05$). YP was also positive significantly correlated to TNF and TNI (table 7).

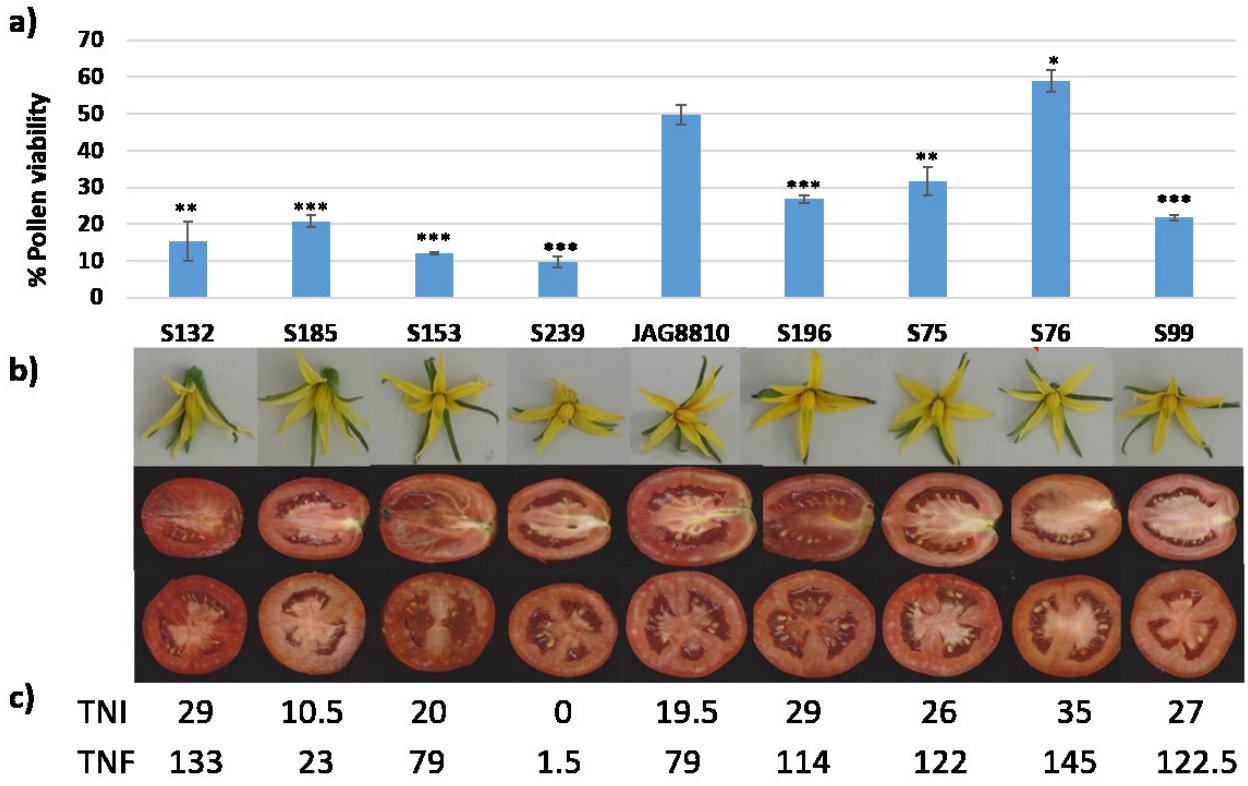



**Figure 6. Comparison among the 8 contrasting JAGF4 genotypes and JAG8810. a) Percentage of Pollen viability, mean ± standard deviation for PV trait. T-test significance level: *, P<0.05; **, P<0.01; ***, P<0.001 is reported. b) flowers, fruit longitudinal sections and fruit transverse sections for each genotype is displayed. c) TNI (total inflorescences number) and TNF (total flowers number) mean value under heat stress conditions are reported.**

| | PV | TNF | TNI | YP |
|---|---|---|---|---|
| **PV** | 1 | | | |
| **TNF** | .443** | 1 | | |
| **TNI** | ns | .961** | 1 | |
| **YP** | .642* | .623** | .602* | 1 |

**Table 7. Correlation analysis among variables tested. PV=pollen viability; TNF: total flowers number; TNI: total inflorescences number; YP: yield production per plant. *P<0.05, **P<0.01, ns= not significant.**

### 2.3.4 Local/systemic effect of heat stress

The main goal of this experiment was to find out if the effect of heat on the treated plant is local or systemic. PV turned out to be unequivocally higher in the part of the plant maintained at RT condition, although the plant is located at the temperature of 33°C. In particular, results indicate that PV of flowers in the bag connected to the pump outside at RT condition (26°C) was 92.8 %, PV of flowers in the bag connected to the air pump in the 33°C room was 9.02% and PV of flowers of plants (without bag) at 33°C was 8.71 %. By contrast flowers of MM plants in the control room at 25°C had PV of 86%. PV of the branch at RT condition results to be 8 times higher than PV of the branch at 33°C and 7 times higher than PV at controlled room. No significant difference was detected between PV of the branch at RT condition and PV at controlled room (figure 7). Moreover, heat stress alterations were observed in young buds and flowers at 33°Cat anthesis stage and were characterized by abnormal anthers and style elongation. Figure 8 shows some phenotypic alterations induced in tomato flowers exposed to prolonged high temperature of 33°C compared to the control.

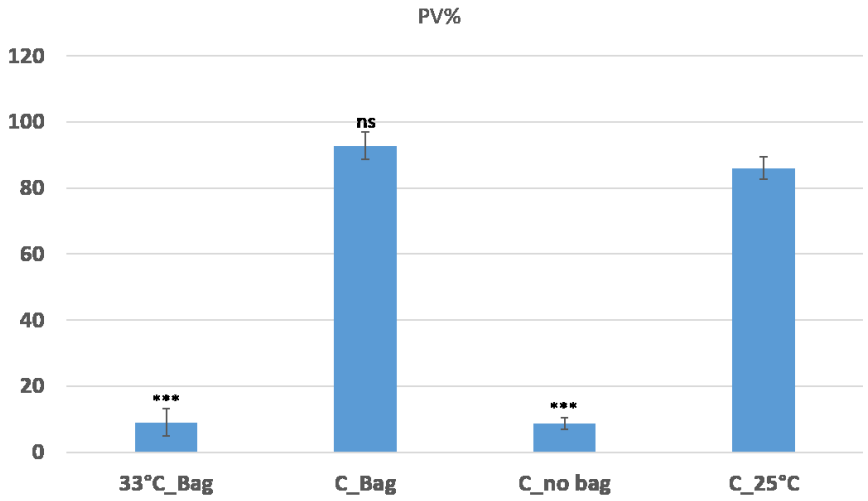



**Figure 7. Pollen viability (PV) comparison between flowers under heat stress condition and control temperatures. Values represent the mean ± standard deviation of PV trait. The differences were assessed by a T-test. ***Significance level: P<0.001; ns=not significant. 33°C_Bag: flowers in the bag connected to the pump in the 33°C room; C_Bag: flowers in the bag connected to the pump outside (temperature range between 26 and 27°C); C_no bag: flowers of plants (without bag) at 33°C; C_25°C: flowers of MM plants in the control room at 25°C.**

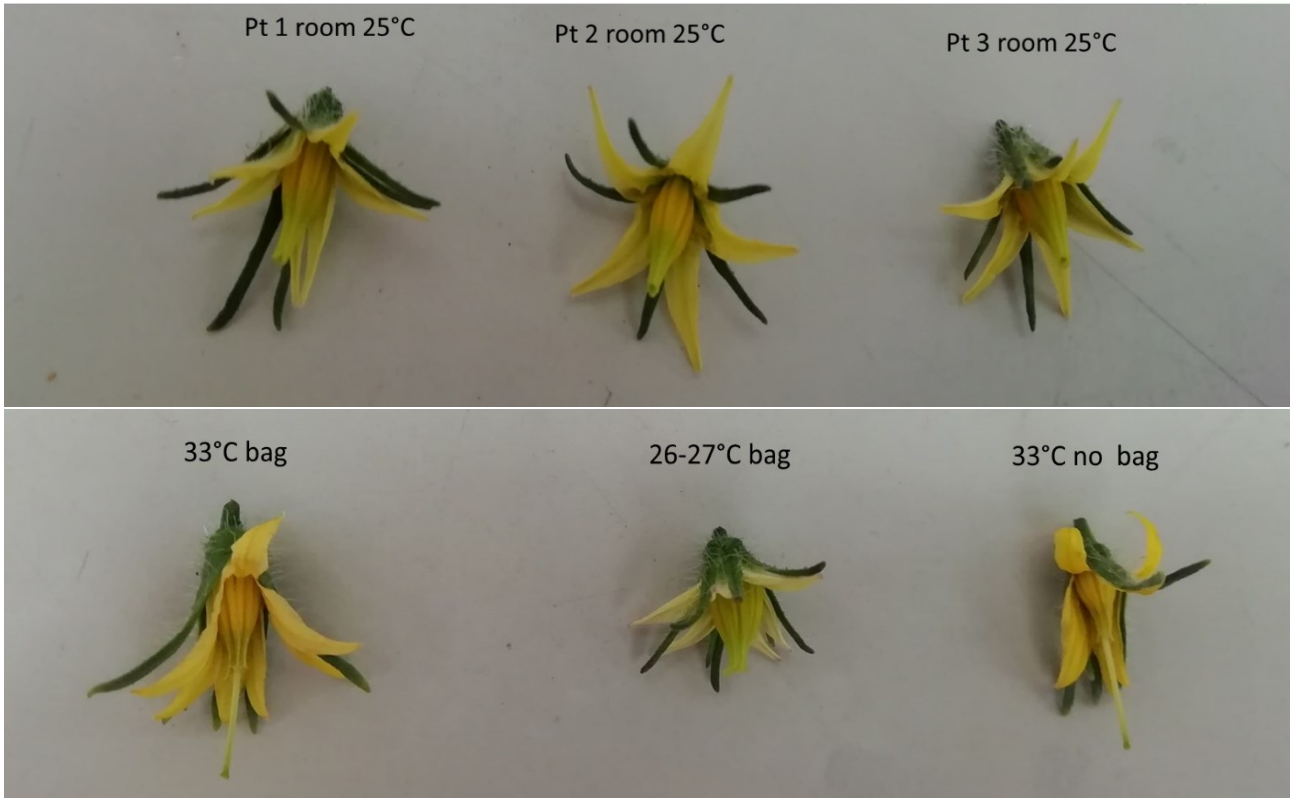



**Figure 8. Phenotypic alterations induced in tomato flowers exposed to prolonged high temperature of 33°C compared to the control. a) flowers of three different plants at 25°C control temperature. b) flowers of plant exposed at 33°C. 33°C_Bag: flowers in the bag connected to the pump in the 33°C room, 26-27 C Bag: flowers in the bag connected to the pump outside; 33°C no bag: flowers of plants (without bag) at 33°C.**

## 2.4 DISCUSSION

Extreme temperature is one of the most harmful abiotic stresses occurring during the plant reproductive phase. In this study, a tomato F4 population was grown imposing a high-temperature condition during flowering and fruit setting in open filed during the summer 2017. Traits related to fruit production such as FS, YP and TFN, which usually are highly affected by high temperatures in tomatoes (Sato et al., 2006; Golam et al., 2012), as well as in other species (Shivanna et al., 1991; Saha et al., 2010), showed a wide variation in our experiment. A strong positive correlation between YP and FS was observed confirming similar findings found in other works (Tele Ayenan et al., 2019). By contrast, a negative correlation between traits related to fruit production such as YP and TFN and fruit quality (SSC) was found. Indeed, it was widely reported that tomato varieties with high SSC tend to be less productive (Garcia et al., 2006; Tigist et al., 2013) although it represents an important tomato selection parameter. The fluctuations in yield productions are governed principally by changes in one or more other components, mainly related to fruit quality (Tigist et al., 2013). However, the

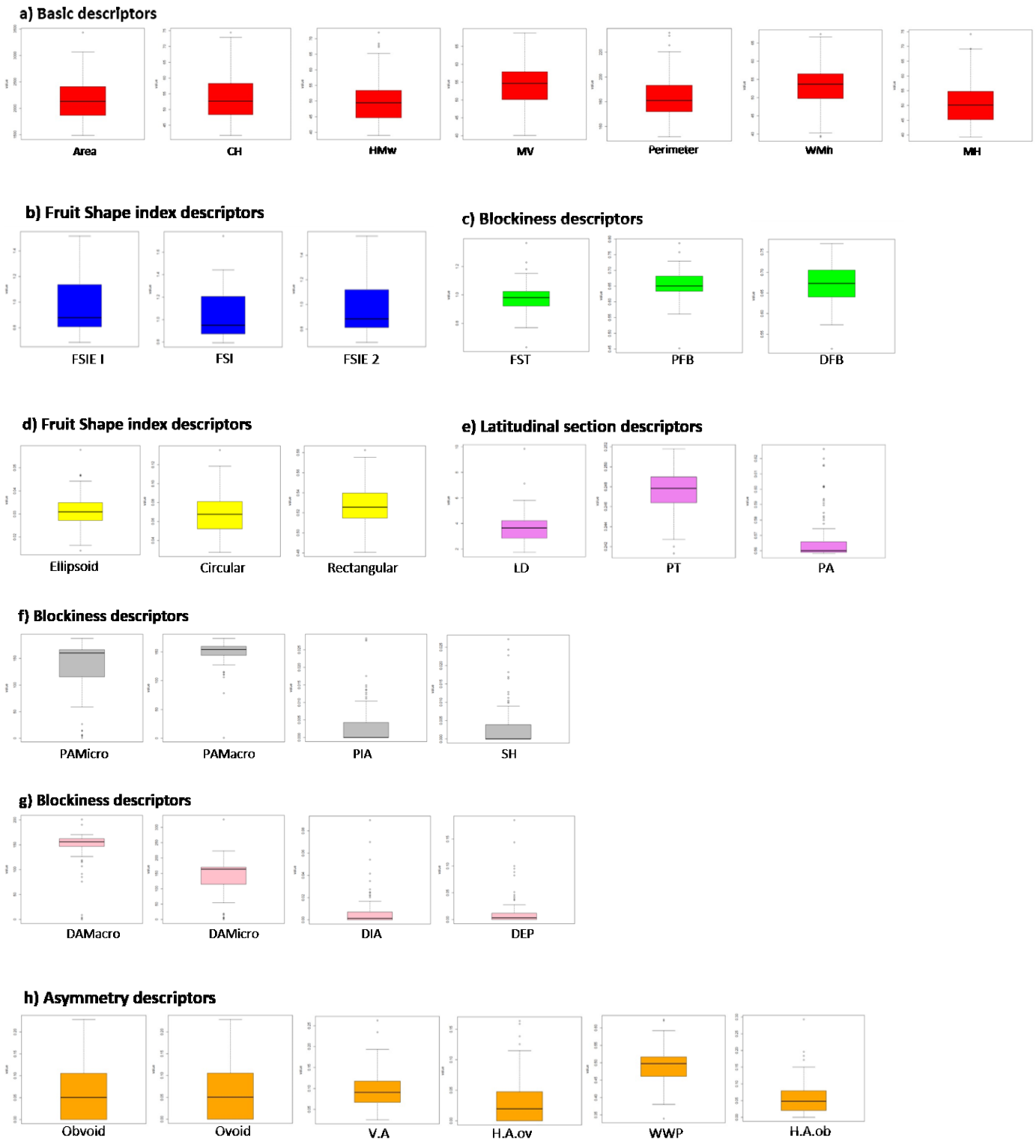
optimal combination of yield and SSC parameters is crucial in tomato breeding since the brix grade represents one of the primary quality traits selected in breeding programs (Fortes et al., 2017).
The PCA analysis used to select the best and worst performing genotypes for yield related traits result very useful for evaluate the contribution to variation of each trait and for identify best lines.
Furthermore, the Tomato Analyzer high-throughput phenomics study allowed the digital acquisition of fruit morphology traits difficult to estimate by the conventional descriptors. This analysis permitted to find an interesting correlation between descriptors related to fruit size and fruit shape in our JAGF4 population. Indeed, several studies have been showed that Tomato Analyzer can be a valuable complementary tool to characterize the commercial varieties and\or to study the genetics of fruit shape in tomato (Mazzucato et al., 2010; Scott, 2010; Panthee et al., 2013; Rodrıguez et al., 2011).
Accurate phenotyping has become a big constraint for multiple traits characterization (D'Agostino and Tripodi, 2017; Araus et al., 2018) and digital phenotyping with scalable technologies can accelerate selection (D'Agostino and Tripodi, 2017). However, the nature of the analyzed traits and the interaction with the external environment must be taken in account. The alignment of phenotyping under controlled conditions with targets for real (i.e., field) phenotyping limit the adoption of new phenotyping tools (Rebetzeke et al., 2016; Daniel et al., 2017). It is very difficult to define how heat stress should be imposed experimentally leading to a high variability in heat treatments and lack of standardized protocols.
To test phenotyping system able to better dissect flowering traits, the best and worst JAGF4 performers, identified through the PCA analysis, were phenotyped for pollen viability through a flow cytometer in controlled conditions (with full control of temperature, light and humidity). Indeed, the tomato pollen viability is reduced under high temperature and the lack of pollen compromises subsequent pollination and fruit production. All environmental conditions were checked since it has been widely demonstrated that also relative humidity affects reproductive processes (Yan et al., 2010; Weerakoon et al., 2008; Driedonks et al., 2018). The pollen viability high-throughput analysis through flow cytometers have several advantages respect the tedious traditional pollen viability assessment made through germination that is important bottleneck in pollen phenotyping (Dreccer et al., 2018). In addition, impedance flow cytometry (IFC) technique used in this thesis is non-destructive, is not species dependent, and can be also used in the field (Heidmann et al., 2016).
Our results confirmed that heat stress had significantly harmful effects on PV (Firon et al., 2006; Pressman et al., 2002; Sato et al., 2006, Xu et al., 2017) although the 4 best performers responded better to heat stress than the worst. Worst genotypes showed PV reduction 8-9 times higher than best performers, which strengthens the notion that pollen viability is correlated with traits related to fruit production (Sato et al., 2000; Firon et al., 2006; Xu et al., 2017). These findings highlight also the

importance of pollen viability as a sub-trait for heat tolerance screening as reported in several crops such soybean, cotton canola, rice, wheat (Kakani et al., 2005; Singh et al., 2008; Mesihovic et al., 2016), confirming its importance as a key indicator of thermo-tolerance status of genotypes. Indeed, the relative simplicity of the sub-traits genetic architecture can widely support indirect assessment of complex fruit setting and yield traits (Paupière et al., 2017; Driedonks et al., 2018). Indirect assessment of heat tolerance based on style length and style protrusion number of flower for plant was also suggested (Xu et al., 2017).

To further investigate the implementation of indirect selection for heat tolerance, we focused also on other two key sub-traits: number of flowers per plant (TNF) and number of inflorescences per plant (TNI). Our finding confirmed that genotypes with high TNI had also higher TNF since they were strongly correlated. A positive significant correlation between TNF and PV in agreement with Xu and collaborators (2017) was also found. Moreover, positive YP significant correlations to PV, TNF, and TNI were highlighted in our dataset. Xu et al (2017) found a positive correlation between fruit set and pollen viability of 0.72, highlighting that PV is a key trait influencing fruit setting and consequently yield under heat stress. Therefore, indirect selection for fruit setting and other traits related to fruit production through PV can be adopted also in field phenotyping.

HSR in plants can be local, when heat effect is only registered in the part of the plant directly exposed to heat, or systemic, if after the stress the heat stress response appears in the entire organism. From data obtained in this work it is evident that heat stress has a tomato plant local response. Moreover, floral morphology showed clearly the damage impaired by heat stress (33˚C) compared to the flowers at control temperature of 25 ˚C. The experiment performed suggested that heat stress affects only the directly exposed parts of the plant, while in the rest of the organism there are no detectable effects. This aspect should be carefully considered when the exposed part is also suffering other stresses (e.g. nutritional deficiency, water scarcity) because they can magnify the heat stress plant response.

## 2.5 SUPPLEMENTARY



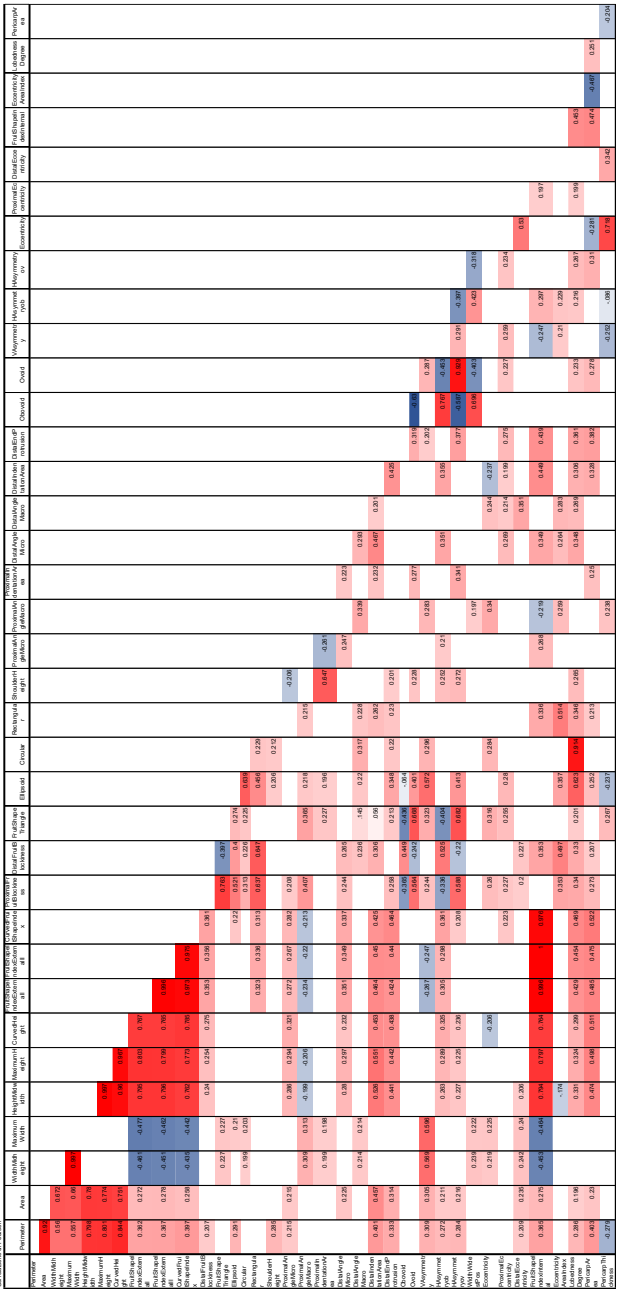

**Supplementary figure 1. Boxplot representation of 38 descriptors based on mean and range of variation observed in JAGF4 population. Each color represents a specific category of Tomato Analyzer descriptors.**

| TRAITS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **LINE** | **YP** | **TFN** | **SSC** | **CR** | **Virosis** | **FS** | **IN** | **LC** | **FRL** |
| **F4S231** | 4.5 | 60 | 4 | 5 | 1 | 42 | 1 | 3 | 1 |
| **F4S150** | 4.8 | 80 | 4.9 | 4 | 1 | 8 | 2 | 3 | 2 |
| **F4S31** | 10.8 | 218 | 4.4 | 1 | 0 | 47 | 1 | 4 | 1 |
| **F4S62** | 8 | 139 | 4.7 | 1 | 0 | 59 | 2 | 4 | 1 |
| **F4S33** | 10.8 | 224 | 3.3 | 3 | 0 | 46.6 | 4 | 3 | 1 |
| **F4S30** | 11.6 | 306 | 3.7 | 1 | 0 | 43.5 | 4 | 5 | 1 |
| **F4S53** | 16 | 350 | 3.3 | 1 | 0 | 53.7 | 3 | 4 | 4 |
| **F4S196** | 12.5 | 249 | 4 | 2 | 0 | 50 | 5 | 3 | 3 |
| **F4S128** | 6.1 | 169 | 4 | 3 | 0 | 33 | 2 | 2 | 1 |
| **F4S4** | 6.3 | 139 | 4.3 | 1 | 0 | 30 | 1 | 4 | 1 |
| **F4S112** | 8.5 | 145 | 4.8 | 3 | 0 | 40 | 3 | 3 | 3 |
| **F4S12** | 4 | 82 | 4.6 | 4 | 1 | 28 | 1 | 1 | 1 |
| **F4S194** | 9.8 | 177 | 4 | 1 | 1 | 57 | 5 | 2 | 2 |
| **F4S34** | 12 | 310 | 4.7 | 5 | 1 | 68 | 3 | 2 | 2 |
| **F4S97** | 6.5 | 127 | 5 | 3 | 1 | 29 | 1 | 3 | 1 |
| **F4S91** | 11.5 | 270 | 3.8 | 1 | 0 | 63 | 3 | 4 | 4 |
| **F4S1** | 7.5 | 129 | 5 | 5 | 1 | 21.75 | 1 | 2 | 3 |
| **F4S82** | 10 | 147 | 4.8 | 3 | 1 | 42 | 1 | 4 | 1 |
| **F4S160** | 5 | 80 | 5 | 3 | 0 | 10 | 4 | 3 | 1 |
| **F4S22** | 9 | 115 | 4.3 | 5 | 1 | 40 | 1 | 2 | 1 |
| **F4S225** | 3.5 | 70 | 4.1 | 5 | 0 | 34 | 2 | 4 | 1 |
| **F4S75** | 13.5 | 330 | 3.6 | 1 | 1 | 73 | 4 | 3 | 3 |
| **F4S100** | 5.6 | 80 | 6.8 | 1 | 1 | 43 | 1 | 4 | 1 |
| **F4S99** | 12 | 208 | 4.1 | 1 | 0 | 65 | 3 | 4 | 5 |
| **S133** | 3 | 125 | 5.5 | 5 | 1 | 29 | 1 | 2 | 1 |
| **F4S13** | 16 | 295 | 4.2 | 1 | 0 | 45 | 5 | 4 | 5 |
| **F4S20** | 7.2 | 150 | 4 | 5 | 1 | 45.25 | 3 | 4 | 4 |
| **F4S113** | 6 | 100 | 5.5 | 3 | 1 | 48 | 1 | 4 | 1 |
| **F4S118** | 9.8 | 191 | 4.1 | 5 | 0 | 44 | 2 | 1 | 2 |
| **F4S8** | 13.7 | 248 | 4.1 | 1 | 0 | 47 | 5 | 4 | 3 |
| **F4S117** | 6.5 | 229 | 5 | 5 | 1 | 66 | 1 | 2 | 2 |
| **F4S2** | 9.7 | 201 | 4.5 | 5 | 1 | 38.25 | 3 | 1 | 4 |
| **F4S41** | 16 | 354 | 4.5 | 1 | 0 | 74.26 | 2 | 5 | 1 |
| **F4S90** | 6 | 132 | 4.7 | 5 | 1 | 35.5 | 1 | 2 | 1 |
| **F4S69** | 10 | 209 | 4.5 | 1 | 0 | 38 | 1 | 5 | 1 |

| TRAITS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **LINE** | **YP** | **TFN** | **SSC** | **CR** | **Virosis** | **FS** | **IN** | **LC** | **FRL** |
| **F4S208** | 2 | 50 | 4.2 | 5 | 1 | 41 | 3 | 3 | 1 |
| **F4S98** | 10 | 215 | 4.3 | 1 | 1 | 54.25 | 3 | 4 | 2 |
| **F4S68** | 9.8 | 152 | 4.3 | 1 | 1 | 35 | 1 | 2 | 3 |
| **F4S204** | 4.5 | 90 | 4.6 | 3 | 1 | 16 | 1 | 4 | 1 |
| **F4S76** | 13.7 | 333 | 5 | 1 | 0 | 53 | 1 | 2 | 1 |
| **F4S28** | 10.5 | 230 | 4 | 1 | 0 | 78 | 4 | 4 | 1 |
| **F4S121** | 7.3 | 180 | 4.6 | 1 | 0 | 35 | 2 | 4 | 1 |
| **F4S187** | 4 | 125 | 4.6 | 1 | 0 | 19.25 | 4 | 3 | 1 |
| **F4S218** | 6.6 | 110 | 4.8 | 1 | 0 | 24 | 3 | 2 | 1 |
| **F4S6** | 5 | 91 | 4 | 1 | 0 | 27 | 1 | 4 | 1 |
| **F4S60** | 13 | 335 | 4.2 | 1 | 0 | 40 | 1 | 5 | 2 |
| **F4S183** | 5 | 60 | 5.8 | 5 | 1 | 20 | 1 | 2 | 1 |
| **F4S224** | 8 | 145 | 4.4 | 1 | 0 | 50 | 4 | 4 | 1 |
| **F4S235** | 4.5 | 50 | 4.2 | 3 | 0 | 20 | 1 | 2 | 1 |
| **F4S156** | 3.4 | 110 | 4.4 | 3 | 0 | 20 | 5 | 5 | 1 |
| **F4S185** | 6.5 | 77 | 4 | 1 | 0 | 24 | 2 | 5 | 1 |
| **F4S244** | 10 | 205 | 3.5 | 1 | 1 | 29 | 3 | 2 | 2 |
| **F4S106** | 9.6 | 167 | 4 | 3 | 0 | 31 | 1 | 4 | 1 |
| **F4S39** | 11.7 | 211 | 4.5 | 1 | 0 | 53.7 | 2 | 3 | 1 |
| **F4S237** | 5 | 55 | 3.2 | 3 | 0 | 22 | 2 | 3 | 3 |
| **F4S249** | 12.5 | 130 | 3.7 | 1 | 0 | 65 | 4 | 2 | 4 |
| **F4S220** | 4 | 58 | 4.7 | 3 | 0 | 15 | 2 | 5 | 2 |
| **F4S180** | 6.7 | 115 | 5.4 | 5 | 1 | 27 | 5 | 2 | 1 |
| **F4S127** | 9.3 | 173 | 5.5 | 1 | 0 | 40 | 2 | 5 | 1 |
| **F4S115** | 4.8 | 119 | 4.5 | 5 | 0 | 60 | 1 | 2 | 1 |
| **F4S205** | 7.2 | 80 | 4 | 5 | 1 | 44 | 1 | 2 | 2 |
| **F4S153** | 2.3 | 45 | 4 | 3 | 1 | 27 | 4 | 2 | 4 |
| **F4S55** | 10.1 | 199 | 3 | 1 | 0 | 32.5 | 4 | 5 | 2 |
| **F4S67** | 8.2 | 145 | 3.9 | 3 | 1 | 30 | 2 | 3 | 1 |
| **F4S36** | 9.7 | 176 | 4.2 | 1 | 0 | 47 | 1 | 4 | 1 |
| **F4S201** | 4 | 105 | 4.6 | 4 | 0 | 24 | 2 | 4 | 1 |

| TRAITS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **LINE** | **YP** | **TFN** | **SSC** | **CR** | **Virosis** | **FS** | **IN** | **LC** | **FRL** |
| **F4S221** | 10.25 | 195 | 5.4 | 5 | 0 | 35 | 2 | 5 | 2 |
| **F4S92** | 15 | 239 | 4.2 | 1 | 0 | 41 | 3 | 5 | 5 |
| **F4S254** | 6 | 100 | 3.8 | 1 | 1 | 22 | 1 | 3 | 3 |
| **F4S47** | 1.65 | 62 | 4.5 | 5 | 1 | 44 | 1 | 4 | 1 |
| **F4S147** | 7 | 106 | 4.5 | 3 | 0 | 33 | 1 | 2 | 3 |
| **F4S84** | 9.5 | 178 | 4.8 | 2 | 1 | 47 | 3 | 3 | 3 |
| **F4S10** | 8.9 | 213 | 5.2 | 1 | 1 | 51.25 | 1 | 4 | 1 |
| **F4S179** | 7 | 100 | 3.8 | 1 | 1 | 25 | 1 | 3 | 3 |
| **F4S5** | 10 | 221 | 3.7 | 1 | 0 | 43.5 | 3 | 4 | 1 |
| **F4S63** | 7.7 | 160 | 3.6 | 1 | 0 | 43 | 2 | 4 | 2 |
| **F4S64** | 8.7 | 200 | 4.7 | 1 | 0 | 20 | 2 | 5 | 1 |
| **F4S159** | 10 | 200 | 3.9 | 3 | 1 | 9 | 2 | 2 | 1 |
| **F4S66** | 5.5 | 70 | 5.2 | 5 | 1 | 35 | 2 | 1 | 1 |
| **F4S126** | 0.8 | 32 | 5.4 | 3 | 0 | 27 | 2 | 5 | 1 |
| **F4S188** | 6.2 | 105 | 4.1 | 1 | 0 | 19 | 2 | 4 | 1 |
| **F4S48** | 13 | 242 | 4.5 | 2 | 1 | 54 | 3 | 2 | 1 |
| **F4S114** | 6 | 115 | 5 | 3 | 1 | 28 | 1 | 4 | 1 |
| **F4S77** | 11.8 | 222 | 4 | 2 | 0 | 47 | 3 | 5 | 1 |
| **F4S51** | 8.7 | 173 | 3.6 | 5 | 0 | 23 | 2 | 3 | 5 |
| **F4S174** | 4.3 | 92 | 4.5 | 5 | 1 | 50 | 3 | 1 | 1 |
| **F4S21** | 8 | 140 | 4.5 | 5 | 0 | 57.2 | 3 | 1 | 2 |
| **F4S139** | 2 | 50 | 4.6 | 5 | 1 | 31 | 1 | 4 | 1 |
| **F4S239** | 4 | 80 | 5 | 5 | 1 | 20 | 3 | 1 | 1 |
| **F4S9** | 4.5 | 184 | 4.5 | 5 | 1 | 41.25 | 3 | 1 | 3 |
| **F4S105** | 11.5 | 246 | 4.7 | 1 | 0 | 34 | 3 | 5 | 1 |
| **F4S104** | 5.8 | 88 | 4.3 | 3 | 0 | 8 | 1 | 5 | 1 |
| **F4S27** | 13.8 | 238 | 4 | 1 | 0 | 61.5 | 5 | 3 | 1 |
| **F4S130** | 7.5 | 90 | 4.1 | 1 | 0 | 51 | 3 | 3 | 1 |
| **F4S78** | 4 | 78 | 4.4 | 4 | 0 | 20 | 2 | 5 | 4 |
| **F4S83** | 5.2 | 148 | 4.5 | 3 | 1 | 47 | 1 | 3 | 1 |
| **F4S252** | 3.5 | 70 | 5.1 | 3 | 1 | 22 | 1 | 4 | 1 |
| **F4S255** | 13 | 235 | 4.7 | 2 | 1 | 45 | 4 | 2 | 1 |
| **F4S132** | 4 | 70 | 5 | 5 | 1 | 15 | 1 | 2 | 1 |

**Supplementary table 1. Mean values of phenotypic traits analysed under heat stress conditions during the summer 2017.**



**Supplementary figure 2. Correlation values between pairs of phenotypes. Only correlation coefficients with P value<0.01 are shown. Colour intensity is proportional to the correlation coefficients. The red indicates positive correlations; the blu indicates negative correlations.**

# CHAPTER III

# GENOME INVESTIGATIONS ON BREEDING LINES CHARACTERIZED FOR HEAT STRESS TOLERANCE

## 3.1 INTRODUCTION

Heat stress is a major abiotic factor limiting the tomato crop productivity since it is particularly sensitive to this stress. In temperate regions, heat stress can become the bottleneck factor of tomato fruit production during hot season.

During the 90s tomato breeders integrated traditional phenotypic assessment with marker-assisted selection (MAS) improving the plant selection process targeting chromosomal region containing quantitative trait locus (QTL) or single trait (Collard and Mackill, 2008; Andolfo et al., 2014; Capuozzo et al., 2017). Several studies have been successfully performed in tomato to map genes or QTLs for abiotic environmental stresses (such as salinity, drought, and heat) and for fruit-related characteristics (Osei et al., 2018). However, QTLs related to complex traits are very difficult to transfer and MAS resulted more suitable for simple traits with a few major-effect genes than for traits controlled from a large number of small effect genes (Dekkers and Hospital, 2002; Heffner et al., 2009). In recent years, the tomato genome sequencing effort and subsequently development of genomic resources and tools, have improved the efficiency and accuracy of tomato genotyping. The potential breeding value of an individual can be estimated using genomic-based data such as single nucleotide polymorphisms (SNP). High-throughput genotyping (HTG) systems generate several thousand SNP markers allowing to scan entire genomes at a reasonable cost.

Genomic selection (GS) provides new opportunities for increasing the efficiency of plant breeding programs for traits with polygenic inheritance (Heffner et al., 2009; Crossa et al., 2010; Lorenz et al., 2011). GS relies on extensive use of molecular markers with inclusion of all marker information in statistical models to estimate genomic breeding value. Most of the variation due to minor QTL is captured in the prediction model. GS can accelerate the genetic gain obtained in each cycle, especially when selection is performed for traits with not highly heritability. Although the effect of each marker is very small, a large amount of genome-wide marker information has the potential to explain all the genetic variance (Wang et al., 2018). The development of statistical methods capable of accurately

predict marker effects has led to the breakthrough of GS increasing the rate of genetic gain per unit of time. GS combines genotypic and phenotypic data from a training population (TRN) in a training set (TRS) to obtain the genomic estimated breeding values (GEBVs) of individuals in a testing population (TST). The GS model is then employed to predict breeding values of not-phenotyped individuals in the next selection step (figure 1).

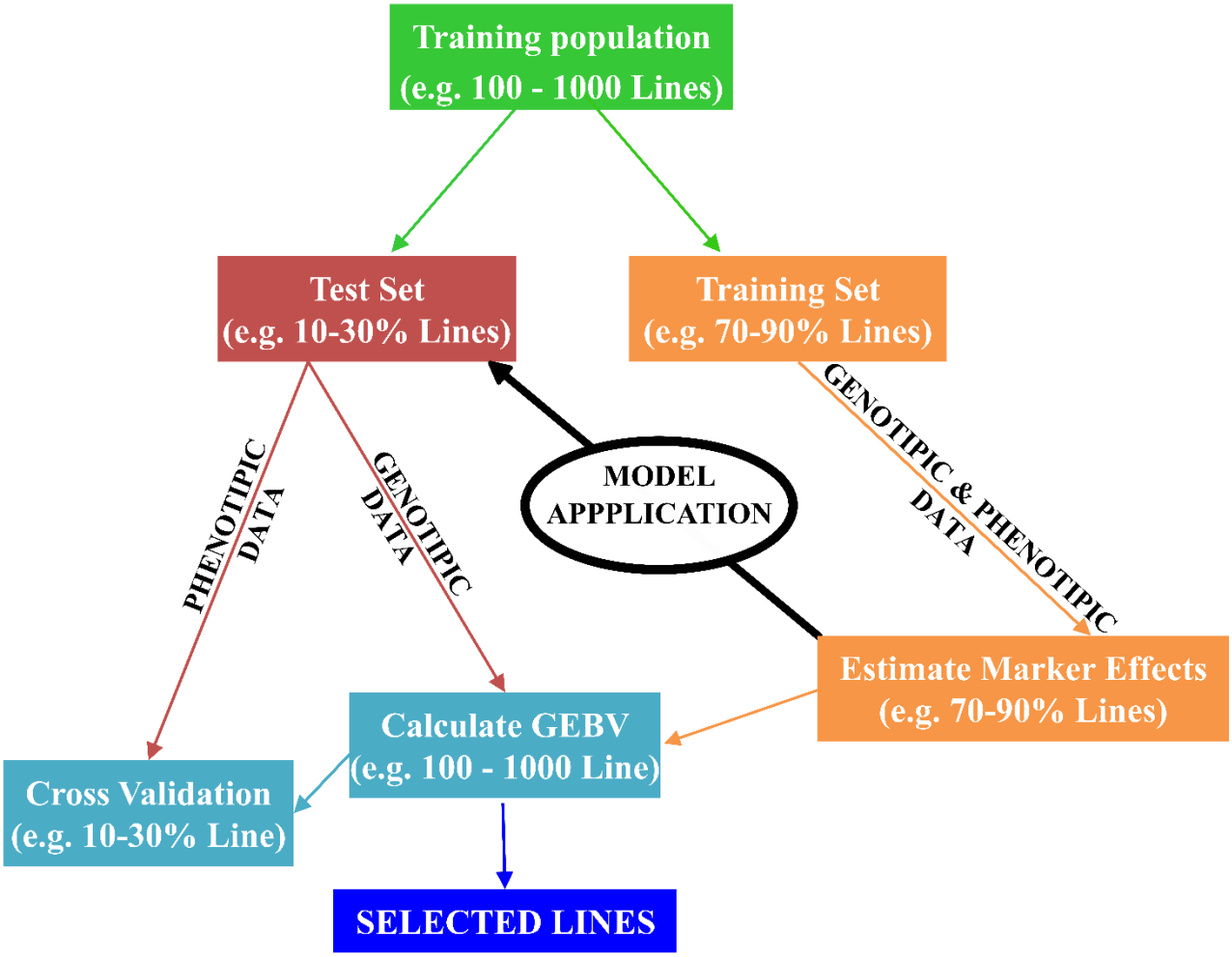



**Figure 1. Genomic selection experimental scheme**

In tomato, pioneer GS studies have been conducted for yield-related traits in fresh market varieties and wild related species (Duangjit et al., 2016; Yamamoto et al., 2016). More recently, Yamamoto et al. (2017) assessed the potential of GS to improve soluble solids content and total fruit weight, analyzing the genome of fresh market F1 tomato varieties and Liabeuf et al. (2018), the resistance to

Bacterial Spot Resistance. The GS models were employed to predict phenotypes of progeny and parents, although the efficiency varied with the parental cross combinations and the selected traits.
GS can result in a valuable tool for tomato breeding even if protocols need to be optimized and validated. Given the fact that many GS programs in this species are starting now, factors affecting the construction, the model accuracy, such as the importance of TRN size, the relationship between individuals in TRS and TST, the marker metrics and the design of GS schema were not yet evaluated and it seems useful to optimize protocols in tomato as well.
In this study, through the use of the GBS (Genotyping by sequencing) approach, we have developed a GS model that can be leveraged to accelerate breeding for heat tolerance in tomato. Furthermore, once the model was built up and validated, the meaningful SNPs (calculated by the model) were used to perform a QTL analysis to better understand the genetic architecture of heat tolerant traits in tomato.

## 3.2 MATERIALS AND METHODS

### 3.2.1 DNA extraction and library preparation and sequencing

DNA was extracted from JAG8810 variety, 100 individuals of the JAGF4 population and 54 individuals of deriving F5 population in 2018 and 2019 respectively, using the Qiagen DNeasy Plant kit. DNA samples were sent to the IGAtech for genotyping by sequencing (GBS) approach.
*In silico* analysis on the reference genome was used to select the best combination of the two restriction enzymes SphI and MboI and the best fragment size distribution to obtain the desired number of loci. Genomic DNA was fluorimetrically quantified, normalized to a uniform concentration and double digested. Libraries to obtain ddRAD were produced using an IGATech custom protocol, with minor modifications respect to Peterson et al. (2012). Fragmented DNA was purified with AMPureXP beads (Agencourt) and ligated to barcoded adapters. Samples were pooled on multiplexing batches and bead purified. For each pool, targeted fragments distribution was collected on BluePippin instrument (Sage Science Inc.). Gel eluted fraction was amplified with oligo primers that introduce TruSeq indexes and subsequently bead purified. The resulting libraries were checked with both Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) and Bioanalyzer DNA assay (Agilent technologies, Santa Clara, CA). Libraries were processed with Illumina cBot for cluster generation on the flowcell, following the manufacturer's instructions and sequenced with V4 chemistry paired end 125bp mode on HiSeq2500 instrument (Illumina, San Diego, CA).

### 3.2.2 Data processing and variant calling

Raw sequencing data were checked for quality score using the BBDuk software (version 35.59), removing low quality portions while preserving the longest high quality part of NGS reads. The minimum length was set at 35 bp and the quality score at 25.

The high quality reads were aligned against the *Solanum lycopersicum* reference genome sequence (SL3.0/) with BWA aligner (Li and Durbin, 2009) (software package for mapping low-divergent sequences against a large reference genome). Default parameters and selection of uniquely aligned reads (i.e. reads with a mapping quality >4) were used. Filtering of detected loci using the Populations program included in Stacks v2.0 (Catchen et al., 2013). Populations was run with option –r=0.75 in order to retain only loci that are represented in at least the 75% of the population.

The "ref_map.pl" program of program belongs to Stacks v2.0 package was used for variant calling. (Catchen et al., 2013). This program aligned data to the reference genome. A VCF file was created and filtered by Minimum Read Depth (DP) >= 4. Variant Calling Filter Files (VCFs) containing all the identified variants were obtained.

### 3.2.3 Circos plot analysis

Four contrasting JAGF4 genotypes (2 tolerants and 2 susceptibles) and the hybrid JAG8810, were analyzed with CircosVCF tool, an interactive tool that provides genomic variation information considering all genetic variant data generated from variant call file (VCF). All parameters were set on the user interface. The karyotype of *Solanum lycopersicum* SL3.0 was loaded in txt format. A line, representing each SNP location, was colored based on its genotype, according to its homozygosity to the reference allele (which in our case is represented by the most frequent allele in the analyzed population), homozygosity to the alternative or heterozygosity. This option was selected for the evaluation since the parental genome was not available (Krzywinski et al., 2009). In the end, also the annotation gene file was loaded to visualize genes harbored by the regions of interest.

### 3.2.4 Genomic selection model construction set up

Following the scheme reported in figure 1, we developed a GS experimental scheme. In particular, 100 phenotyped and genotyped JAG F4 individuals were used to build up a training population. The TRN was divided in TRS e TST in a cross-validation scheme, meaning that the model was build using X individuals of the TRS and validated in the TST with remaining 100-X lines. The GS model was

then employed to predict the genetic breeding values (GEBVs) of not-phenotyped individuals in the next selection step (F5) and validated in F6 offspring deriving form F5 population through a phenotypic characterization under heat stress conditions.

### 3.2.5 GS statistical model

The rrBLUP model (Endelman, 2011) was used for analyzing our SNP dataset. It measures the whole set of marker effects at one time under the assumption that genomic variances for all loci are equal based on the following relation:

$$y = \mu + Xg + e$$

where y is Nx1 vector of phenotypic means; μ the overall mean of the training set; X is the NxNm (marker matrix); g is NmX1 (marker effects matrix) and e represents the Nx1 vector of residual effects. Matrixes were constructed numerically coding markers as 1, -1, 0, and NA respectively for major allele, minor allele, heterozygosity, and missing data. Loci with missing data were imputed using the EM method with the A.mat function of the 'rrBLUP' package in R (Endelman, 2011). A.mat also calculates the additive relationship matrix.

### 3.2.6 TRNs composition and marker dataset selection

Cross-validation of prediction accuracy was conducted varying the training set (TRS) size for each trait. In particular for YP, where the TRN used to build up the model was composed from 90 individuals, 15, 20, 25, 30 and 35 individuals were randomly included in the TST to predict their GEBVs and replicated 1000 times. While for SSC all individuals were included in the analysis (100) and 20, 25, 30, 35, 40, 45 individuals were randomly included in the TST.

To evaluate the effects of GBS marker selection on genomic prediction accuracy, two criteria for filtering SNPs were considered. GBS SNPs were filtered for minor-allele frequency (MAF) >0.05 and based on 5 levels of percentage of eliminated missing values (PEMVs). These 5 levels were 90, 85, 80, 75, 70 %. Markers were filtered based on all possible combinations of PEMV and MAF producing 5 GBS marker subsets (e.g., 5 PEMV levels × 1 MAF level). After filtering, remaining missing values were imputed using the EM method with the A.mat function of the 'rrBLUP' package in R (Endelman, 2011).

These 5 subsets were used to evaluate the impact of the TRS on prediction accuracy. Therefore, these 5 marker dataset were combined with the 5 TRS dataset described above both for YP and for SSC.

In particular, the accuracies of predictions, obtained by the relationship between predicted and observed values, was assessed by Pearson correlation setting up 25 rrBLUP models, one for each combination. The prediction accuracy (Pearson correlation coefficient between the GEBVs and the phenotypes of the TST) of each combination was evaluated using cross validation replicated 1000 times. The mean predictive ability across the 1000 replicates over the folds was used as the elimination criterion. The validated models obtained were then used to predict F5 individuals in the following cycle (which have been only genotyped but not phenotyped) (see section 3.2.1).

### 3.2.7 JAGF6 population phenotyping for GS validation

The F6 offspring deriving from JAGF5 population was grown and phenotyped as described for JAGF4 population in chapter II-section 2.2.2, except for the planting binary pattern.

Briefly, 54 F6 lines were grown in a completely randomized design with 10 plants per line. Seeds were sown in plateau under plastic-house and seedlings were transplanted in field under plastic tunnel in the first half of May in order to expose growing plants to high temperatures.

Tomato plants were grown following the standard cultural practices of the area and temperature. At fruit red ripe stage, the Soluble Solid Content (SSC-3 fruits per plant) and yield production per plant (YP) were measured in order to validate the model.

### 3.2.8 QTL analysis

Quantitative trait locus (QTL) analyses for the two traits (YP and SSC) assessed in JAGF4 population (chapter II) were performed using R-based software package, R/qtl (Broman et al., 2003; Broman and Sen, 2009, in R version 3.2.5). From the dataset used for GS model construction (14.286 for SSC and 14.210 for YP) a total of 2.486 and 2.034 SNP were used for mapping QTL in YP and SSC respectively.

After filtering, SNP markers were coded as AA, BB, AB, and NA respectively for major allele, minor allele, heterozygosity, and missing data. To analyze the F4 population the function cross <- convert2bcsft (cross, BC.gen = 0, F.gen=4) was used. Genetic maps were constructed by est.map using the Kosambi's function of R-qtl (Kosambi, 1944). Phenotypic traits were analyzed by composite interval mapping (CIM) using the maximum likelihood via the expectation-maximization algorithm (Dempster et al., 1977). Using 1 cM steps, CIM was calculated using default settings. Genome-wide significance thresholds were generated for each trait by determining LOD values at alpha of 0.05 from 1000 permutations (Broman and Sen,2009). The percent phenotypic variance explained by a QTL was estimated via the following formula: h2=1-10 − (2/n) LOD, where "n" is the

sample size and “LOD” the single-QTL model LOD score, using single interval mapping (SIM). The whole genome was scanned in steps of 1 cM by the scan one function. Batch number was added as covariate in both analyses, whereas family number was only applied as covariate in the interspecific mapping population. The 95% Bayes credible interval was assessed using the bayesint function, using the LOD score of CIM.

### 3.2.9 Variant annotation and investigation of candidate genes

In order to perform a variant annotation of the Variant Calling Files (VCF) (see section 3.2.2), the SnpEff software was employed on JAGF4 population dataset, which let the association of each variant to the annotated genes and to predict their effect on the protein function. Variants were classified based on their location (e.g., exon, intron, splice site, etc.) and the effect of their ‘impact’: High, Moderate, Low and Modifier. A blastn analysis of candidate genes against TAIR database (https://www.arabidopsis.org) was performed to further investigate their function in the model species *A. thaliana*.

## 3.3 RESULTS

### 3.3.1 High-throughput genotyping of tomato segregating populations

The hybrid variety JAG8810 and the segregating populations JAGF4 and JAGF5 have been sequenced through GBS approach in 2018 and 2019 respectively, producing a total of 117.798.988 and 136.926.433 reads.

After quality check, all reads were mapped onto the reference genome (*Solanum lycopersicum* cv. Heinz 1706), and the results were used for SNP calling using STACKS package (Catchen et al., 2013) producing 135.355 SNPs in JAGF4 and 176.596 SNPs in JAGF5. In order to characterize the JAGF4 population genetic diversity, further analyses on identified variants were performed. The total number of SNPs for each chromosome is displayed in figure 2. A high density of SNPs on chromosomes 4,5 6 and 11 was found.

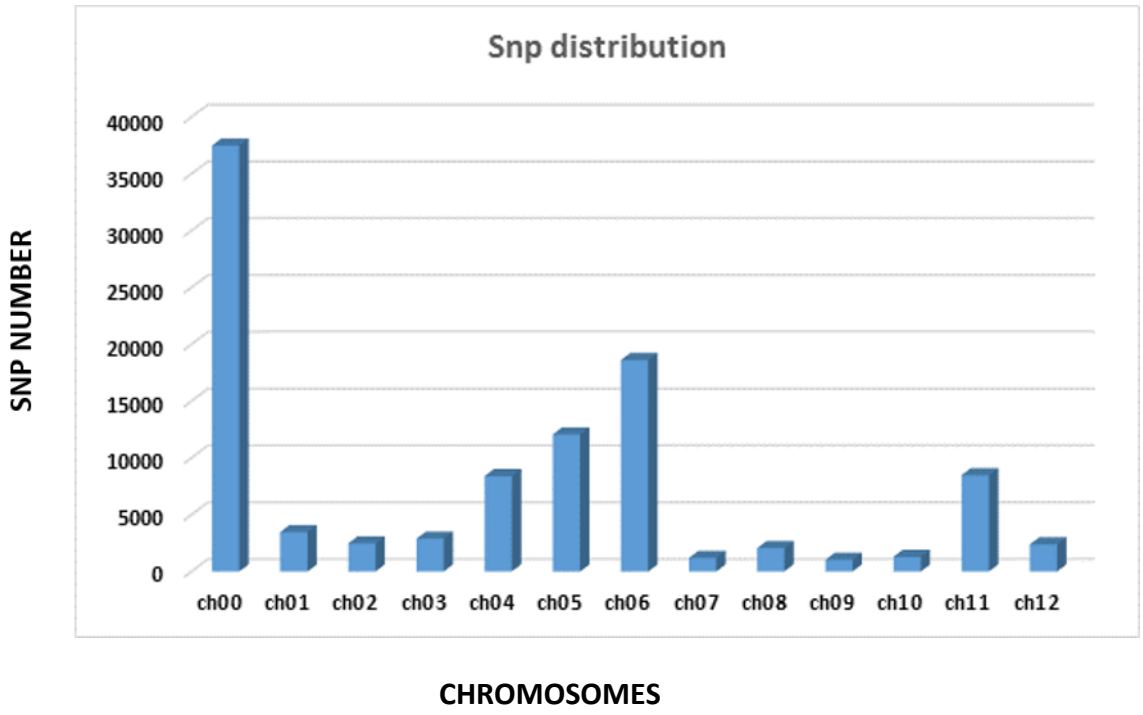



**Figure 2. SNP number distribution among chromosomes.**

Based on these results, to better investigate the genomic relatedness of individuals based on whole genome SNP information and to detect specific meaningful SNP regions on these chromosomes associated with traits of interest, a circos plot analysis was obtained. For this analysis we used 4

contrasting genotypes in terms of traits involved in fruit production (YP-Yield production per plant, % FS-percentage of fruit set, TFN-total fruit number): 2 tolerant (S77 and S75) and 2 susceptible (S160 and S239) lines using the hybrid JAG8810 as reference genotype, were compared. Results were shown in figure 3. The outermost circle depicts the tomato chromosomes, whereas the five inner circles reports, from the outside to the inside, the number of SNPs found in S239, S160, JAG8810, S75 and S77 genotypes, respectively. SNPs homozygous to the reference allele are colored in yellow, the alternative form in red and heterozygous SNPs in blue (table 2). Clearly differentiating regions can be detected at a glimpse of the generated circos plot between the contrasting genotypes also compared with JAG8810 (figure 3a). In particular, results highlight two specific contrasting chromosomic regions: one on chromosome 4 and the other on chromosome 6. To better investigate such regions, the SNPs found on two chromosomes of interest and related genes annotation, were used to generate a new plot (figure 3b). The results clearly showed that hybrid JAG8810 revealed a heterozygous state, whilst genotypes with higher value in terms of fruits production (YP, % FS and TFN) (table 1), represented by the 2 intermost cycles, are homozygous for the A allele (dominant allele) and genotypes with lower values (the two outermost cycles) are homozygous for the B allele.

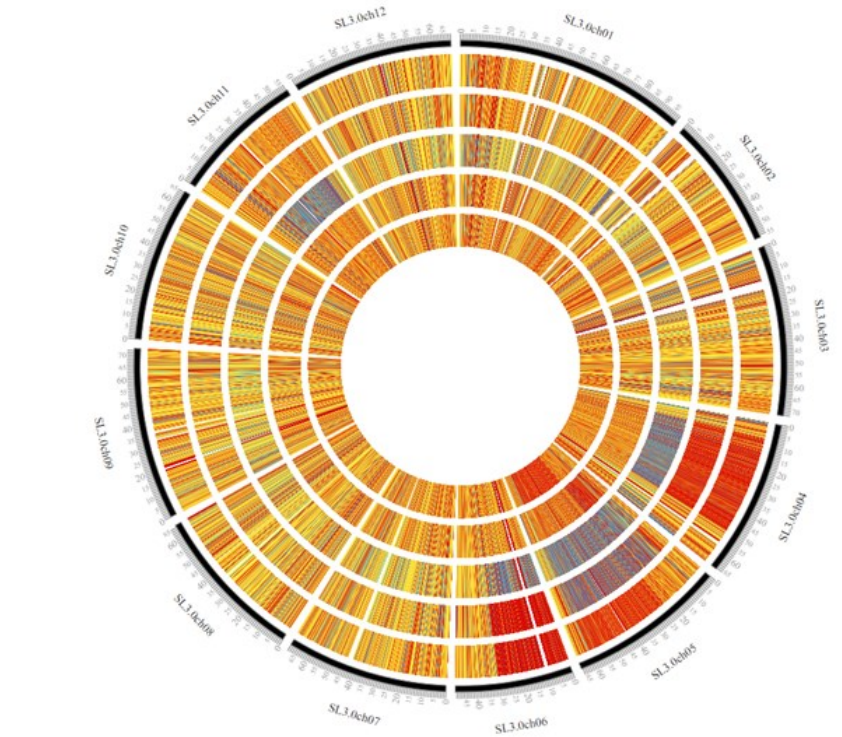
| | Genotype | YP(kg) | FS% | TFN(num) | SSC |
|---|---|---|---|---|---|
| Worst | S239 | 4.0 | 20 | 80 | 5.0 |
| | S160 | 5.0 | 10 | 80 | 5.0 |
| | S77 | 12.5 | 55 | 248 | 3.9 |
| Best | S75 | 13.5 | 73 | 330 | 3.8 |

**Table 1. Phenotypic values of YP, %FS, TFN and SSC for S239, S160, S77, S75 genotypes. The phenotypic values is refer to the specific line used for DNA extraction.**

| | Statistics | Base Pairs per Line | SNPs per Line |
|---|---|---|---|
| S239 | **Homozygous to Reference:** 5889 SNPs<br>**Heterozygous:** 1647 SNPs<br>**Homozygous to Alternative:** 24856 SNPs<br>**Total:** 32392 SNPs | 1 bp |  <br>Homozygous to Reference<br>Heterozygous<br>Homozygous to Alternative |
| S160 | **Homozygous to Reference:** 5579 SNPs<br>**Heterozygous:** 1475 SNPs<br>**Homozygous to Alternative:** 25338 SNPs<br>**Total:** 32392 SNPs | 1 bp |  <br>Homozygous to Reference<br>Heterozygous<br>Homozygous to Alternative |
| JAG8810 | **Homozygous to Reference:** 7440 SNPs<br>**Heterozygous:** 10162 SNPs<br>**Homozygous to Alternative:** 14790 SNPs<br>**Total:** 32392 SNPs | 1 bp |  <br>Homozygous to Reference<br>Heterozygous<br>Homozygous to Alternative |
| S75 | **Homozygous to Reference:** 14525 SNPs<br>**Heterozygous:** 1712 SNPs<br>**Homozygous to Alternative:** 16155 SNPs<br>**Total:** 32392 SNPs | 1 bp |  <br>Homozygous to Reference<br>Heterozygous<br>Homozygous to Alternative |
| S77 | **Homozygous to Reference:** 13760 SNPs<br>**Heterozygous:** 1153 SNPs<br>**Homozygous to Alternative:** 17479 SNPs<br>**Total:** 32392 SNPs | 1 bp |  <br>Homozygous to Reference<br>Heterozygous<br>Homozygous to Alternative |

**Table 2. Color scheme and number of SNPs related to homozygous to reference, heterozygous and homozygous to alternative, were reported for S239, S160, JAG8810, S75 and S77 individuals.**

a)




b)

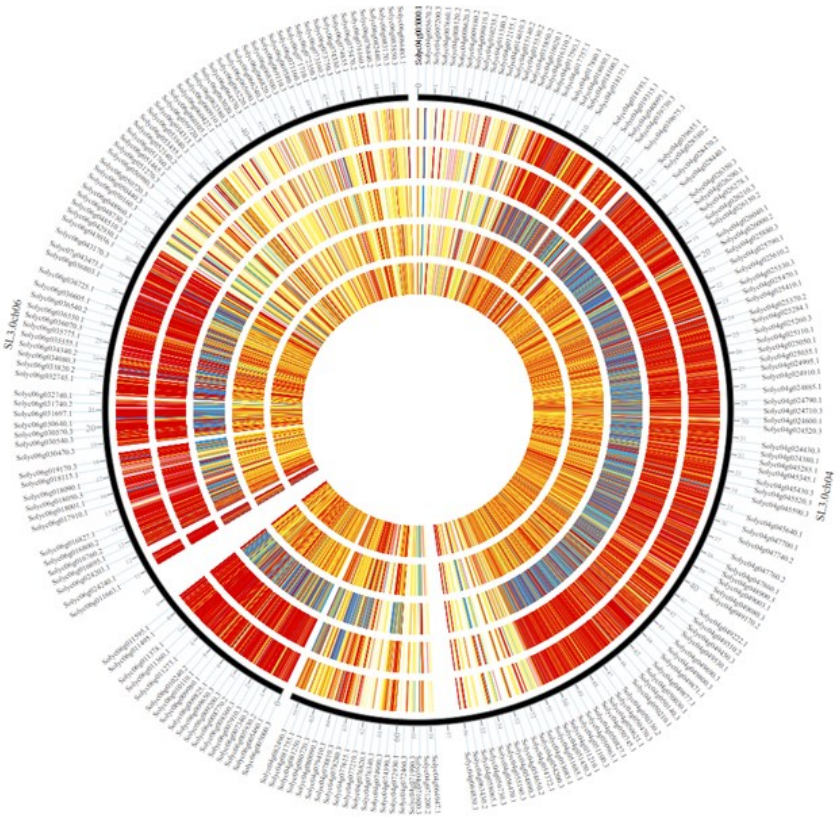



**Figure 3. Circos plots representation. a) The outermost circle displays the 12 tomato chromosomes whereas the five inner circles represent, from outside inside, the SNP's genotypes of S239, S160,**

**JAG8810, S75 and S77 respectively. Each line is colored according to the SNP's genotype, where yellow stands for homozygosity to the reference allele, red for homozygosity to the alternative, and blue heterozygosity. b) Focus on chromosome 4 and 6. In this representation also the genes annotation was displayed.**

As widely established in chapter II, traits related to fruit production such as YP, FRN, % FS were negatively correlated to the fruit quality trait SSC. To better investigate this correlation also at genotypic level, values relates to SSC were also reported in table 1 highlighting the negative correlation. The chromosomic regions involved in SSC under heat stress conditions were the same of those related to fruit production (chromosome 4 and 6) but with opposite effect. Genotypes with lower SSC value, represented by the 2 intermost cycles, are homozygous for the A alleles (dominant allele); whereas genotypes with higher values (the two outermost cycle) are homozygous for the B alleles. These results, mirroring the negative correlations found between the phenotypic traits themselves analyzed in chapter II, suggesting that traits involved in YP and SSC under heat stress conditions are located on chromosome 4 and 6.

### 3.3.2 Development and optimization of genomic selection models for tomato crop

#### 3.3.2.1 GS model experiment set up

GS scheme was implemented following a JAG8810 variety Single Seed Descendent (SSD) selection (figure 4). In particular, after two SSD cycles (F2 and F3), in F4 generation we selected 100 individuals to both genotype by GBS approach and phenotype under heat stress conditions (as described in chapter II). The genotypic and phenotypic data obtained in this step were used as training set to build up and cross-validate a statistical model (MODEL CONTRUCTION), able to predict the performance of not phenotyped F5 lines (MODEL APPLICATION). In last step, the F6 offspring deriving from F5 generation was phenotyped under heat stress condition to validate the prediction of the model (MODEL VALIDATION). The two predicted traits in this experiment were YP and SSC, two critical traits in heat stress condition as widely discussed before (chapter II).

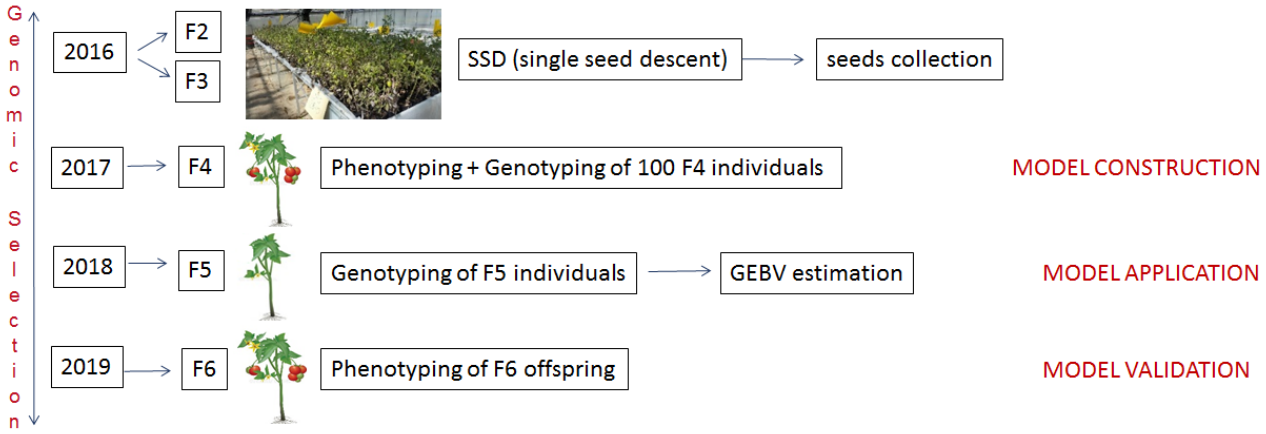



**Figure 4. GS experiment set up**

### 3.3.2.2 Model construction

In this section, the effect of Training Set (TRS) size and the selection of marker subsets used for GS model training was investigated. For each trait, five marker subsets were obtained filtering SNPs for MAF>0.05 and PEMV at following 5 levels: 90, 85, 80, 75, 70 %. Since the number of individuals of TRN was different for the two traits analyzed (90 for YP and 100 for SSC) these subsets included different number of SNPs (table 3).

| PEMV | Number of markers | |
|---|---|---|
| | SSC | YP |
| 90 | 14.286 | 14.210 |
| 85 | 16.625 | 16.224 |
| 80 | 18.806 | 18.637 |
| 75 | 21.019 | 20.555 |
| 70 | 23.471 | 23.228 |

**Table 3. SNP number obtained after filtering by MAF and PEMV for SSC and YP respectively**

Starting with the maximum TRS set of 80 and 70 sampled genotypes for SSC and YP respectively, we randomly removed 5 genotypes, reducing the TRS samples set and evaluating the model quality for each subset. In particular, 5 TRS subsets were created for each trait: 80, 75, 70, 65, 60 (out of 100) for SSC data and 75, 70, 65, 60, 55 (out of 90) for YP. Subsequently, 25 RR-blup models, varying

the combination of two parameters, were tested in order to assess the effect of marker subsets and TRS sizes on prediction accuracy, with 1000 cycles of permutation for model (table 4). The prediction accuracy both for YP and SSC resulted different in each combination tested. As expected, accuracy of the GS models generally decreased when fewer genotypes were used in TRS. Moreover, the results revealed that predictive ability did steadily increase up until the maximum PEMV in the cross validation strategy for both traits and, indeed, it was maximum for the highest PEMV (90%). The mean predictive ability across 1000 replicates over the folds was used as the elimination criterion. At the end, the best prediction was obtained by the model deriving from the combination: MAF>0.05, PEMV (90%) and TRS\TST composition in 80\20 ratio for SSC and MAF>0.05, PEMV (90%) and TRS\TST composition in 75\15 ratio for YP (figure 5).

a)

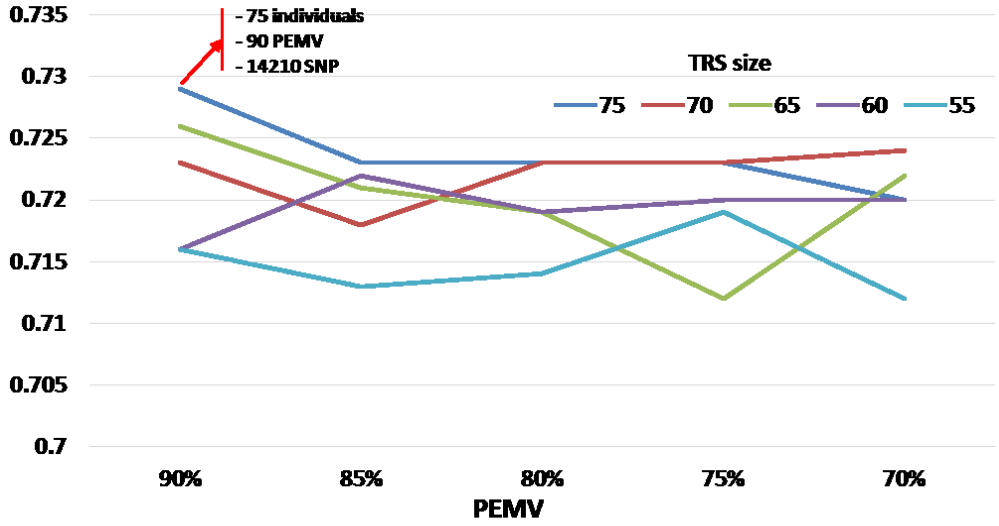
| TRS size | Prediction accuracies of TST GEBVs under different proportion of PEMV | | | | |
|---|---|---|---|---|---|
| | 90% | 85% | 80% | 75% | 70% |
| 80 | 0.715 | 0.708 | 0.705 | 0.705 | 0.711 |
| 75 | 0.711 | 0.71 | 0.708 | 0.71 | 0.709 |
| 70 | 0.71 | 0.709 | 0.708 | 0.711 | 0.701 |
| 65 | 0.704 | 0.707 | 0.704 | 0.707 | 0.707 |
| 60 | 0.705 | 0.705 | 0.705 | 0.704 | 0.702 |

b)

| TRS size | Prediction accuracies of TST GEBVs under different proportion of PEMV | | | | |
|---|---|---|---|---|---|
| | 90% | 85% | 80% | 75% | 70% |
| 75 | 0.729 | 0.723 | 0.723 | 0.723 | 0.72 |
| 70 | 0.723 | 0.718 | 0.723 | 0.723 | 0.724 |
| 65 | 0.726 | 0.721 | 0.719 | 0.712 | 0.722 |
| 60 | 0.716 | 0.722 | 0.719 | 0.72 | 0.72 |
| 55 | 0.716 | 0.71 | 0.714 | 0.719 | 0.712 |

**Table 4. Prediction accuracies of TST GEBVs under different proportion of PEMVs. a) Mean value of prediction accuracy obtained for SSC after 1000 cycles of iteration. b) Mean value of prediction accuracy obtained for YP after 1000 cycle of iteration.**

a)




b)

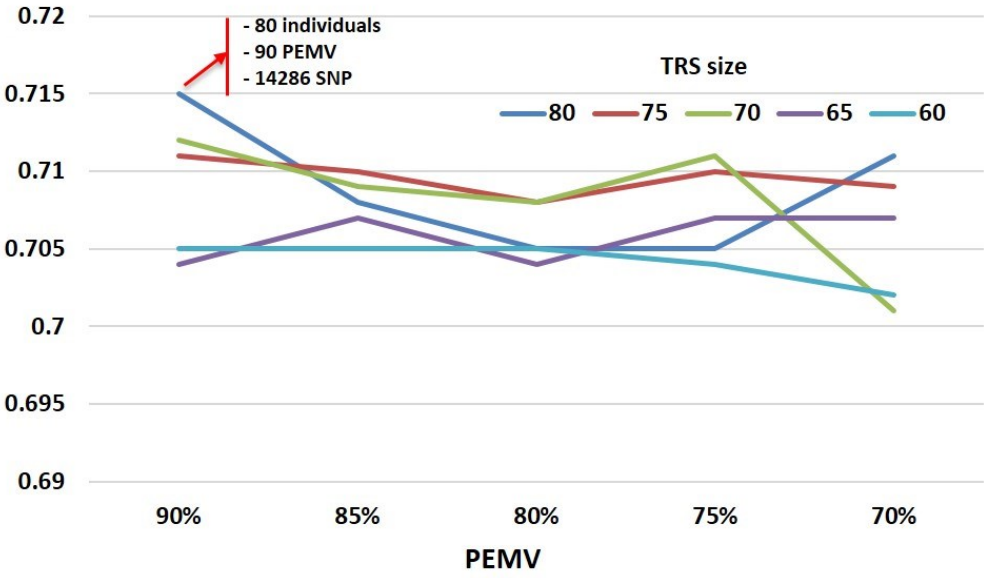



**Figure 5. Prediction accuracies of TST GEBVs under different proportion of PEMVs. The red arrow indicates the best combination for YP (5a) and SSC (5b).**

This result highlights that using higher PEMV values the accuracy of GS model based on larger TRS improves. The marker subsets for the optimized GS model turned out to be of 14.286 and 14.210 SNPs and that represents also the smallest subsets in terms of marker density (table 3). This is an important outcome of our investigation since the use of fewer markers result in significant cost reductions for genotyping, but also impact the extent of LD picked up by the prediction models. Figure 6 and 7 showed the correlation values between GEBVs and real data for each of the 1000 iteration cycles. Extremely accurate models were obtained for YP (0.729) and SSC (0.715) using parameters optimized in our test. This means that we found a good correspondence between the GEBVs and real values in the TST as shown in figure 8 for YP trait.

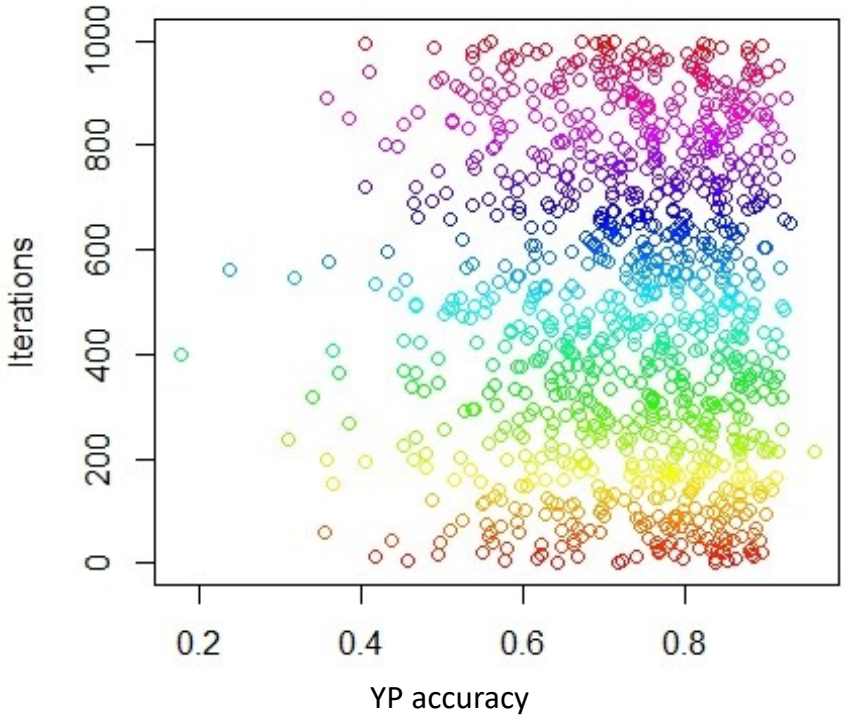



**Figure 6. Plot between correlation values and 1000 cycles of iteration for YP**

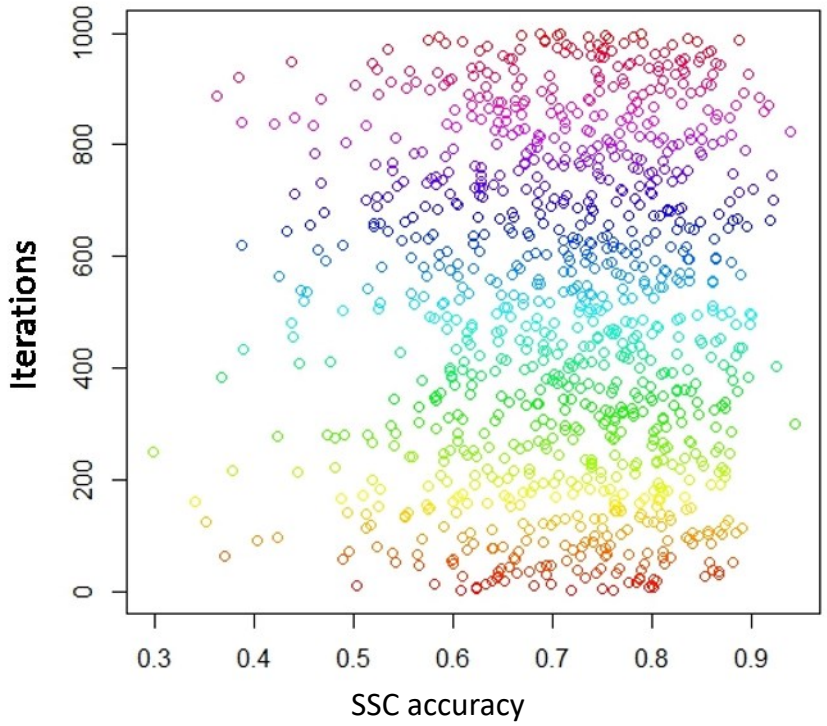



**Figure 7. Plot between correlation values and 1000 cycles of iteration for SSC**

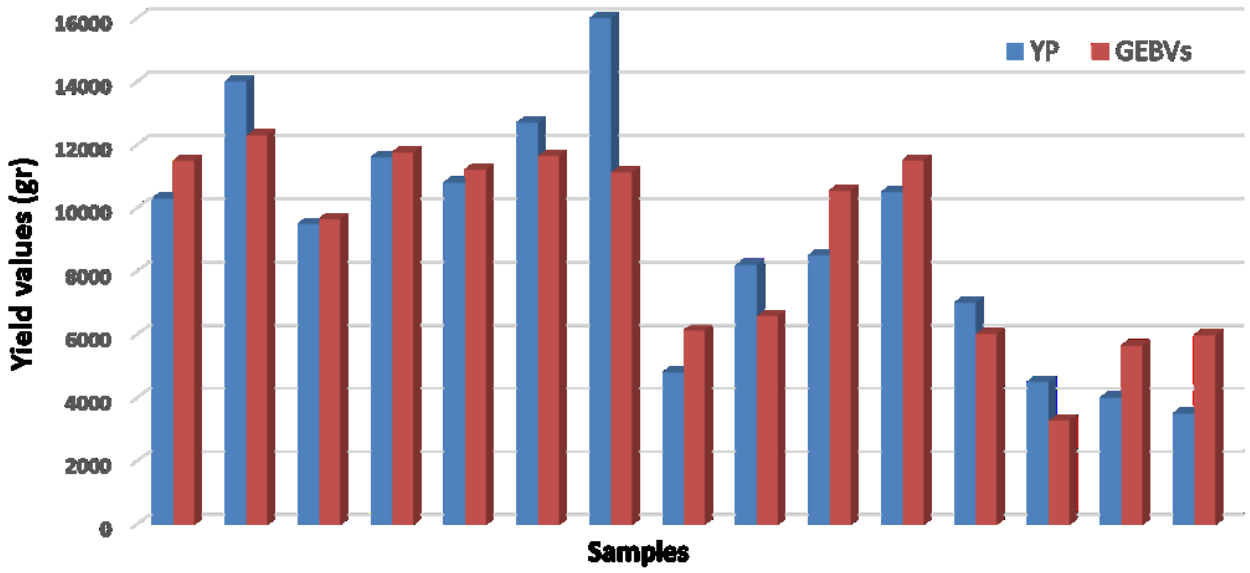



**Figure 8. Comparison between real and predicted value (YP and GEBV) in a TST cross validation.**

### 3.3.2.3 Model application

The model constructed and cross-validated in JAGF4 population was applied in next selection round to estimate GEBVs of F5 progeny population (JAGF5). In particular, 54 JAGF5 individuals were only genotyped by GBS producing a total of 176.596 SNPs. Based on the high-quality SNPs subset selected for model construction in JAGF4, 12.566 shared SNPs were detected among the two dataset. However, 10.648 SNPs were retrieved and used in the Model application since 1.908 SNP showed inconsistencies in the two datasets and therefore, have been eliminated. The GEBVs of JAGF5 individuals (not phenotyped) were calculated by using the GS models trained on the JAGF4 individuals. In particular, the JAGF4 individuals were used as TRS to calculate the GEBVs of the 54 JAGF5 genotypes used as TST. Table 5 summarizes the GEBV predictions obtained in this selection round. High variability both for YP and SSC was observed. Indeed, the GEBVs ranged from a minimum value of 4.1 to a maximum value of 13 with an average of 8.4 for YP (table 5a) and from a minimum value of 3.6 to a maximum of 5.4 with an average of 4.52 for SSC (table 5b). To select the best and worst predicted performers the mean value of all F5 genotypes was used as threshold (table 5). Individuals with GEBVs higher than average value were considered the best predicted performers and individuals with lower, the worst. Figure 9 displays GEBVs of JAGF5 individuals for

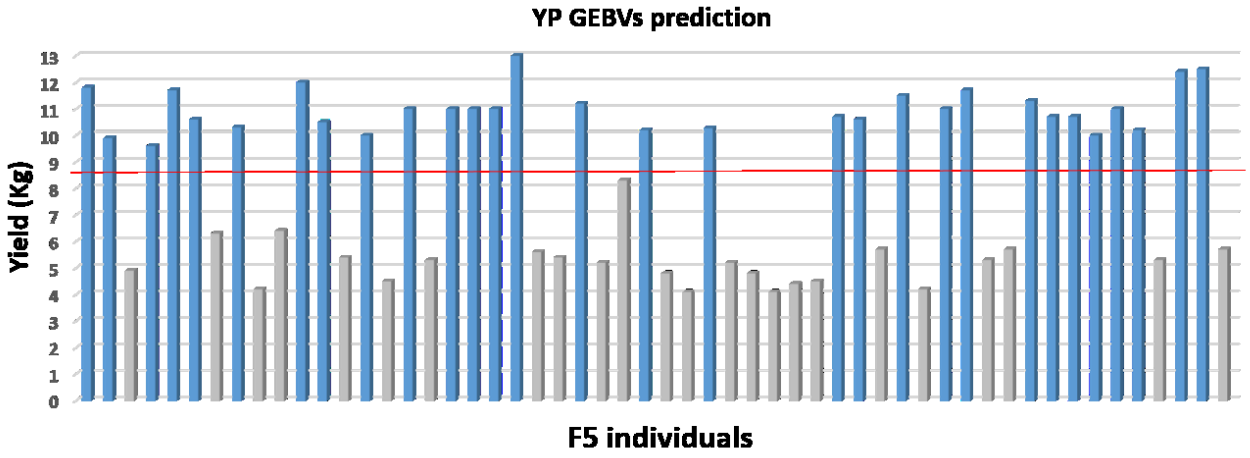
the two traits. For YP the 30 best predicted performers, with higher GEBV than 8.4, are colored in blu (figure 9a) whereas for SSC, the 25 best genotypes (with GEBV higher than 4.52) are colored in green (figure 9b).

| Predicted Yield | |
|---|---|
| Mean (kg) | 8.4 |
| Minimum value (kg) | 4.1 |
| Maximum value (kg) | 13 |
| % Best | 55.5 |
| % Worst | 45.5 |

| Predicted SSC | |
|---|---|
| Mean | 4.52 |
| Minimum value | 3.7 |
| Maximum value | 5.4 |
| % Best | 46.3 |
| % Worst | 53.7 |

**Table 5. The mean, minimum and maximum values of predicted GEBVs in F5 population for YP (a) and SSC (b) were reported. The % of the best and worst predicted values were also reported.**

a)




b)

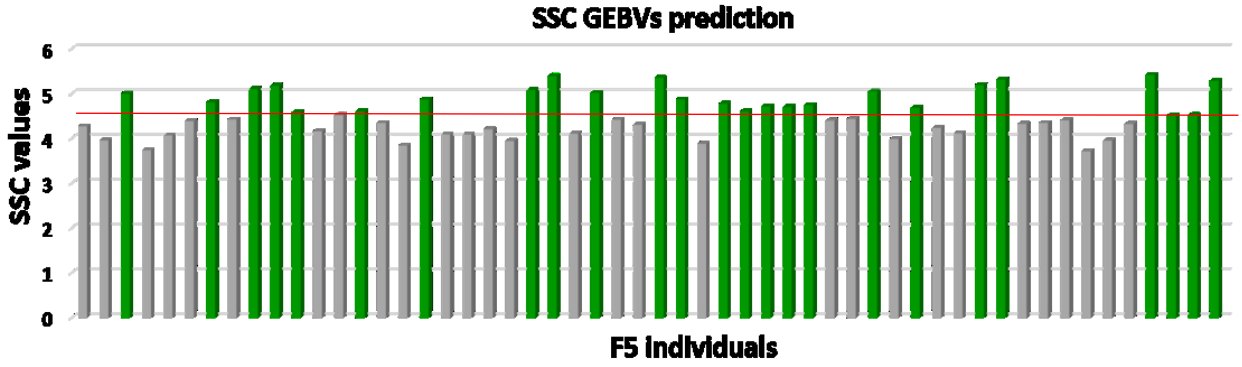



**Figure 9. GEBVs prediction for YP and SSC. a) the best predicted performers for yield are displayed in blu. b) the best predicted performers for SCC are displayed in green. Red lines indicate the threshold values.**

### 3.3.2.4 Model validation

In the last step we investigated whether the GS models efficiently predicted phenotypes in the next progeny populations. For this purpose, the F6 offspring deriving from F5 generation was phenotyped under heat stress condition to validate the prediction of the model at phenotypic level for both analyzed traits (YP and SSC). The coefficients of correlation between the GEBVs calculated on F5 genomic data and the phenotypic values obtained in F6 offspring were comparable to values obtained in F4 model build up (0.729 and 0.715), although it turned out to be slighter lower (0.67 and 0.7 for YP and SSC respectively). The reduction (approx. 4.000 SNPs) of the marker dataset used for the F4 model construction during the F5 model application stage may be affected the prediction precision. Furthermore, the GS model for SSC showed higher predictability than YP probably due to use of the high-precision digital refractometer. However, the results indicated that our GS model was correctly

cross-validate by experimental trials, confirming that the calibrated GS models for YP and SSC set up in this study can efficiently predict phenotypes.

Subsequently, we investigated in more detail the predicted data in order to identify the most interesting lines. Table 6 reported the mean, the minimum and maximum values obtained in F6 population and the percentage of the best and worst performers using the mean value as threshold.

24 individuals had higher yield value than 2.29 (mean value) and therefore represents the best performers in terms of YP, whereas for SSC, 32 genotypes have higher GEBV than 4.54 (mean value) and represent the best. Out of 54 F6 individuals, 41 resulted correctly predicted (74%) for YP and 42 individuals for SSC (76%) (table 6). The performance of best predicted JAGF5 individuals (for the two analyzed traits) was confirmed by the JAGF6 phenotypic validation (table 7).

a)

| YP F6 individuals | |
|---|---|
| Mean (kg) | 2.29 |
| Minimum value (kg) | 0.2 |
| Maximum value (kg) | 6 |
| % Best | 44.5 |
| % Worst | 55.5 |
| % of correctly predicted lines | 74 |

b)

| SSC F6 individuals | |
|---|---|
| Mean | 4.54 |
| Minimum value | 3.6 |
| Maximum value | 5.4 |
| % Best | 59.3 |
| % Worst | 40.7 |
| % of correctly predicted lines | 76 |

**Table 6. The mean, minimum and maximum values of F6 individuals for YP (a) and SSC (b). The percentage of the best and worst performers were also reported.**

| Best Lines | Yiel production | | SSC | |
|---|---|---|---|---|
| | GEBV JAGF5 | YP JAGF6(Kg) | GEBV JAGF5 | SSC JAGF6 |
| S2 | 11.8 | 6 | 4.51 | 4.56 |
| S22 | 9.6 | 4.6 | 5.4 | 5.1 |
| S34 | 10.3 | 2.5 | 4.69 | 4.6 |
| S67 | 10 | 3.1 | 5.18 | 5.2 |
| S77 | 11 | 2.9 | 4.86 | 5.3 |
| S118 | 10.27 | 2.4 | 4.7 | 4.7 |
| S153 | 11 | 4.9 | 4.5 | 4.6 |
| S159 | 11.7 | 3.1 | 5.3 | 5.3 |
| S194 | 10 | 2.9 | 5.05 | 5 |

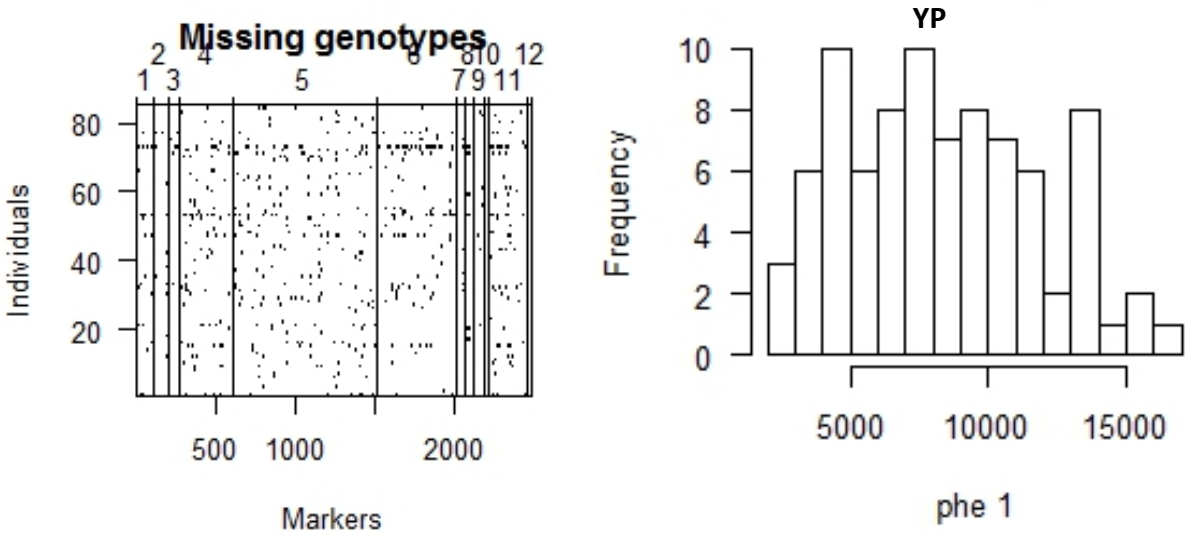
**Table 7. The best predicted performers (for the two analyzed traits) in JAGF5 population. For YP, individuals with GEBVs higher than 8.4 (the mean value in JAGF5) and higher than 2.29 (the mean**

**value in JAGF6) were reported. For SSC, individuals with GEBVs higher than 4.52 (the mean value in JAGF5) and 4.54 (the mean value in JAGF6) were reported.**

### 3.3.3 Mapping of quantitative trait loci associated to heat stress tolerance

A subset of SNPs with higher effect in the rrBLUP model from the datasets of 14.286 and 14.210 SNP was selected to identify specific chromosomic regions associated with the two analyzed traits of interest (YP and SSC) under heat stress condition. In particular, SNP with effect higher than |2| were selected obtaining a total of 2.486 and 2.034 for YP and SSC respectively, covering more than 95% of genome in both cases. Figure 10 reports a summary plot of the phenotypic and genotypic data for two traits.

a)



b)

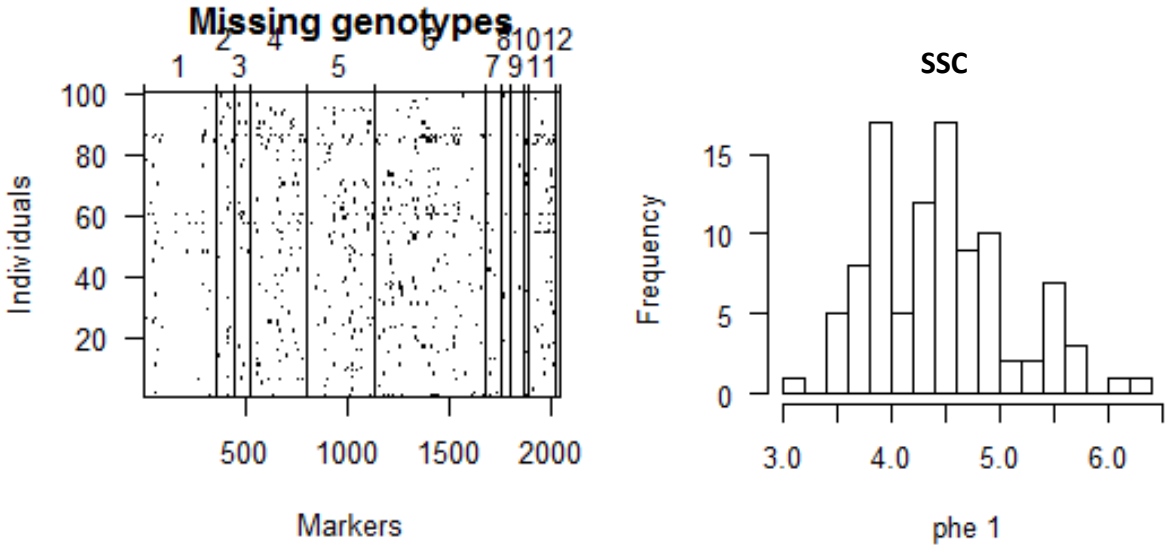



**Fig.**

**Figure 10. Summary plot of the phenotypic and genotypic data including the pattern of missing genotype data (black pixels indicate missing data) and histograms of the two phenotypes distribution for YP (a) and SSC (b).**

Combination of genotyping and phenotyping data of JAGF4 population were analyzed by composite interval mapping (CIM), revealing QTLs for the two investigated traits. In particular, to declare the significant QTLs, a permutation test was performed for each phenotype. The value of the LOD Score threshold for the two traits was 3.00. In this way, 5 significant QTLs were detected for YP, on chromosome 5, 6, 8, 9 and 11 (qYP1, qYP2, qYP3, qYP4, qYP4) explaining the 51, 64, 53, 51, 53 % of the phenotypic variance respectively (Xiao et al., 2011). Whereas only 1 significant QTL was detected for SSC on chr 6 (qSSC1) explaining 48% of the phenotypic variance. The results of the QTL LOD scores values were displayed in figure 11. The highest LOD Scores for the two analyzed traits were found on the same chromosome (chr6) and it was 19.99 for qYP2 at V81972 marker and 12.16 for qSSC1 at V82503 marker.

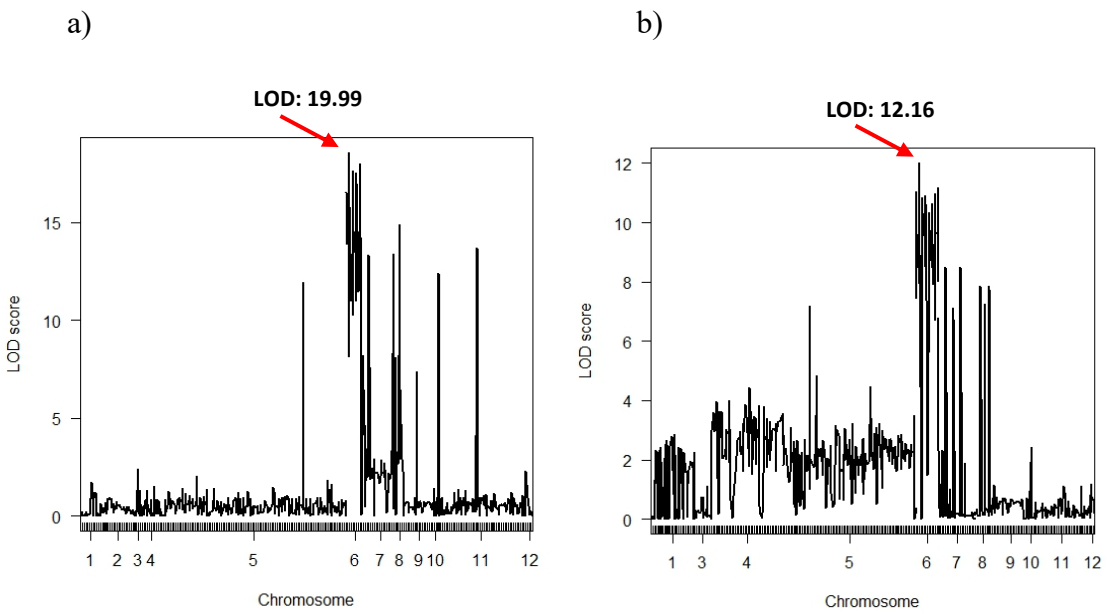



**Figure 11. LOD scores values representation for each chromosomes for YP (a) and SSC(b). The arrows indicated the maximum LOD scores calculated**

Interestingly, QTLs for the two different traits mapped approximately to the same genomic region on this chromosome. In particular, the qYP2 starts at 355.596 bp and ends at 28.961.753 bp whereas qSSC1 starts at 19.772.865 bp and ends at 30.508.757 bp indicating an overlapping co-localization of 9.188.888 bp (figure 12).

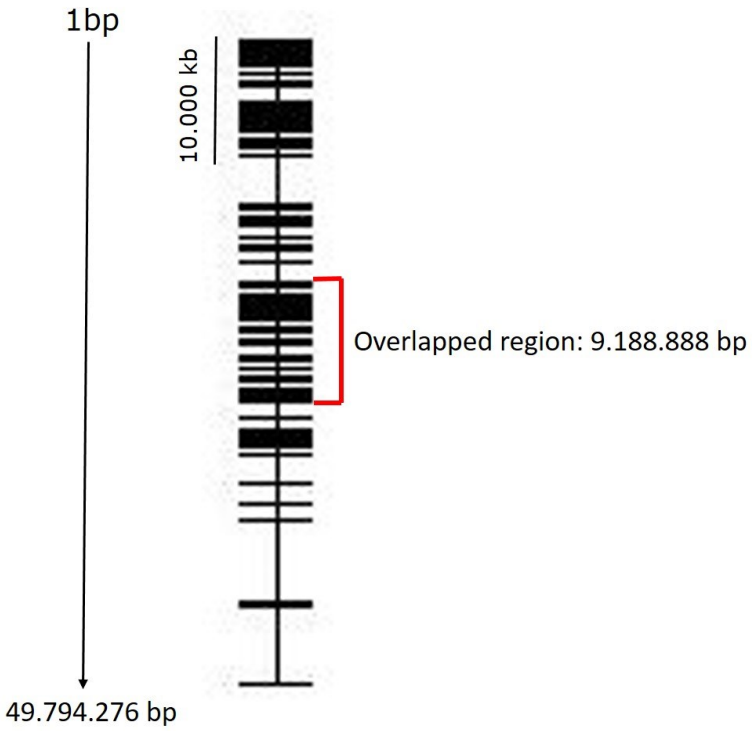



**Figure 12. SNP distribution on chr 6. In red, the overlapped region between qYP2 and qSSC1 has been highlighted**

When the YP and SSC phenotypes were plotted against the genotype at V81972 and V82503 markers, in both instances the heterozygous revealed an intermediate state in terms of YP and SSC but opposite homozygous state was observed. Indeed, for YP, genotypes which are homozygous for the A alleles (dominant allele) showed higher yield while the genotypes homozygous for the B alleles (recessive allele) showed lower yield values. Whereas, for SSC we found a higher SSC degree in homozygous genotypes for the B alleles and lower SSC in homozygous genotypes for the A alleles (figure 13). These results highlight that the co-localization for the two traits on chr 6 perfectly mirrored the negative correlations found between the YP and SCC phenotypic traits (chapter II) confirming they have a genetic basis.

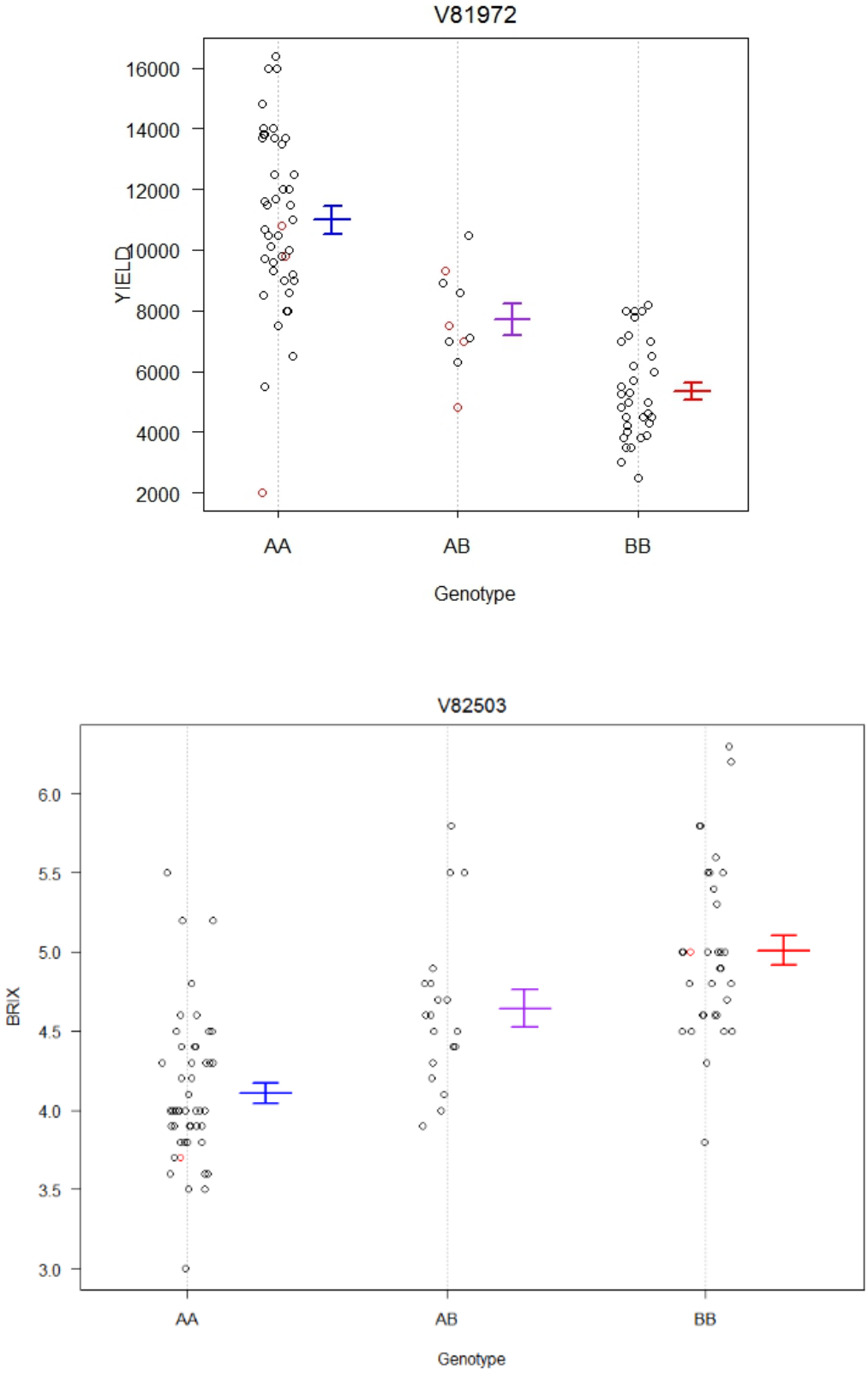
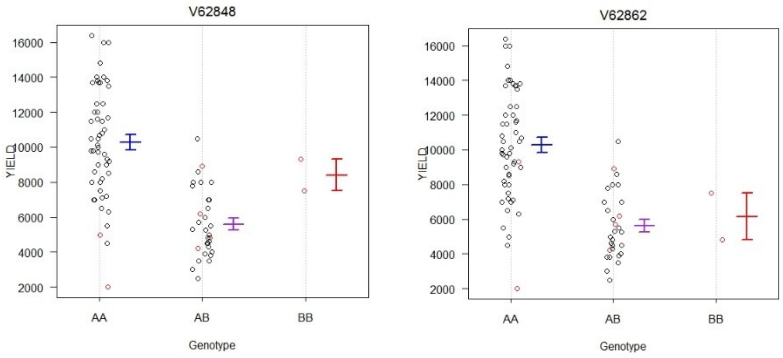



**Figure 13. a) YP phenotype is plotted against the genotype at markers V81972 on chr 6. b) SSC phenotype is plotted against the genotype at marker V82503 on chr 6.**

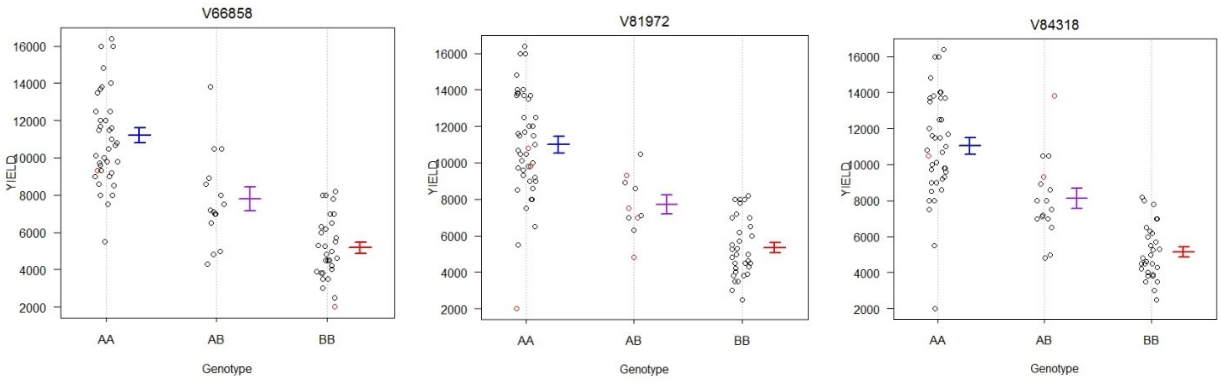
To better investigate the response of plants to heat stress, QTL regions associated with traits of interest were scanned to identify genes putatively involved in the heat stress response.

Five QTLs were identified for the YP under heat stress condition around markers V62848 on chr 5, V81972 on chr 6, V88039 on chr 8, V89258 on chr 9 and V91295 on chr 11. To estimate the location of QTLs an interval characterized by 3 indexes was delimitated: the first and last indicated the ends of the intervals; the middle index corresponds to the location with higher LOD score value. All significant markers with higher LOD score values located on the 5 QTL intervals were plotted again trait of interest (figure 14).

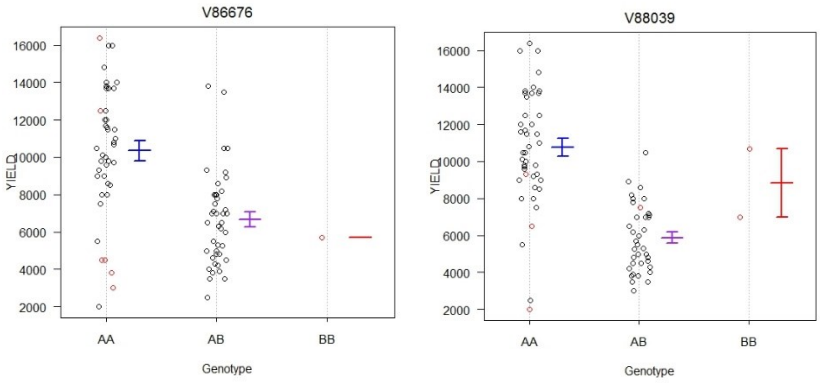
a)



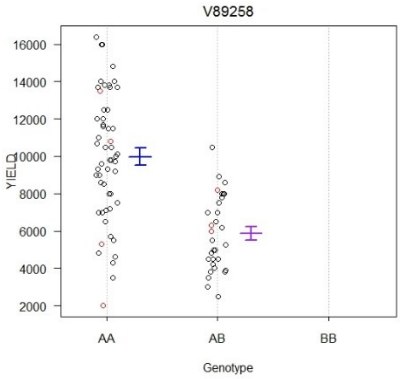

b)



c)




d)




e)

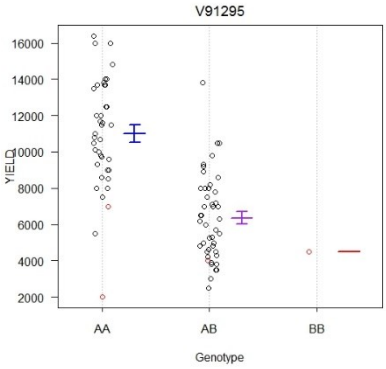



**Figure 14. Plot of all significant markers with higher LOD score values identified as indexes of QTL in 5 chromosome regions again the trait of interest. a) Plot of V62848 and V62862 markers (chr 5); b) plot of V66858, V81972 and V84318 markers (chr 6); c) plot of V86676 and V88039 markers (chr 8); d) plot of V89258 marker (chr 9); plot of V91295 marker (chr11).**

Interestingly, all the three indexes defining the QTL interval on chr 6 correspond to three markers given as input. At this point, in order to identify genes putatively involved in YP under heat stress, a variant annotation was performed to associate each variant included in the identified QTL to the annotated genes and to predict their effect on the protein function. However, we excluded from this analysis all SNPs located in intergenic regions, including all the SNPs found on chr 5 and 9. Table 8 reported the SNP ID, the gene position and gene annotation. The SNPs previously selected as indexes of QTLs intervals are highlighted in red being those with a higher LOD score and putatively more involved in the plant's response to stress. In this way, a SNP located on chr 8, in the intronic region of a stress responsive alpha-beta barrel domain was identified. Interestingly, among all the identified SNPs in the QTL region on chr 6, those that delimited the ends of interval (V66858 and V84318) were located on upstream region of the R2R3 trascription factor myb 59 and the SKP1-like protein respectively.

| CHR | SNP ID | AFFECTED REGION | GENE ANNOTATION |
|---|---|---|---|
| 8 | V86676 | intron | Stress responsive alpha-beta barrel domain protein |
| 6 | V66858 | upstream | R2R3 Trascrition factor myb 59 |
| 6 | V66874 | downstream | Aspartic proteinase-like protein 1 |
| 6 | V66898 | upstream | Pentatricopeptide repeat superfamily protein |
| 6 | V66998 | missense | Leucine-rich receptor-like protein kinase family protein |
| 6 | V67015 | 5'utr | NAD(P)H-ubiquinone oxidoreductase A1 |
| 6 | V67025 | upstream | Retrovirus-related Pol polyprotein from transposon TNT 1-94 |
| 6 | V67081 | 5'utr | Cytochrome b-c1 complex subunit 8 |
| 6 | V67122 | 3'utr | Gibberellin-regulated family protein |
| 6 | V67367 | Missense and synonimous | CW-type zinc finger protein |
| 6 | V67578 | downstream | Pentatricopeptide repeat-containing protein family |
| 6 | V67617 | Downstream and intron | R2R3MYB transcription factor 24 |
| 6 | V67688 | intron | Vps51/Vps67 family (components of vesicular transport) protein |
| 6 | V68227 | upstream | Retrovirus-related Pol polyprotein from transposon TNT 1-94 |
| 6 | V70636 | upstream | Plasma membrane ATPase |
| 6 | V70657 | intron | Chaperone protein |
| 6 | V79825 | upstream | PLAC8 family protein |
| 6 | V80267 | intron | Reverse transcriptase |
| 6 | V81248 | upstream | Retrovirus-related Pol polyprotein from transposon TNT 1-94 |
| 6 | V81292 | downstream | Retrovirus-related Pol polyprotein from transposon TNT 1-94 |
| 6 | V81918 | downstream | Serine palmitoyltransferase |
| 6 | V82916 | intron | Autophagy-related protein 3 |
| 6 | V83212 | intron | MAP kinase kinase kinase 37 |
| 6 | V83269 | downstream | Retrovirus-related Pol polyprotein from transposon TNT 1-94 |
| 6 | V83382 | Miss and synon | Kinase family protein |
| 6 | V83948 | downstream | CASP-like protein |
| 6 | V84318 | upstream | SKP1-like protein |

**Table 8. SNP variants annotation. The chromosome, SNP ID, the affected region and gene annotation were reported. SNPs selected as indexes of QTLs intervals were highlighted in red**

Regarding the SSC only one significant QTL was detected. A number of SNPs located on genes mapping on chromosome 6 QTL interval is reported in table 9.

| CHR | SNP ID | AFFECTED REGION | GENE ANNOTATION |
|---|---|---|---|
| 6 | V81920 | downstream | Serine palmitoyltransferase |
| 6 | V82790 | 5'utr | Dehydration responsive element-binding factor protein |
| 6 | V82909 | upstream | B3 domain-containing protein |
| 6 | V82918 | intron | E3 ubiquitin protein ligase RIN2 |
| 6 | V82990 | downstream | DNA-directed RNA polymerase subunit beta |
| 6 | V83000 | upstream | DNA-directed RNA polymerase subunit beta |
| 6 | V83150 | downstream | Remorin family protei |
| 6 | V83199 | intron | DNA-directed RNA polymerase III subunit RPC3 |
| 6 | V83286 | upstream | Heat shock protein 90 |
| 6 | V83390 | 3'utr | ABC transporter C family protein |
| 6 | V84313 | upstream | SKP1-like protein |

**Table 9. SNP variants annotation. The chromosome, SNP ID, the affected region and gene annotation were reported.**

## 3.4 DISCUSSION

Tomato crop is particularly sensitive to heat stress, a major abiotic factor limiting the crop productivity. By the end of the 21st century (2081–2100), global temperatures are supposed to increase on average of 1 to 3.7 ◦C compared to their levels in 1986–2005 (Pachauri e al., 2014). Therefore, the genetic architecture of some traits associated with heat tolerance (e.g., pollen viability, inflorescence number, fruit set) has been dissected and associated quantitative trait loci (QTL) mapped (Xu et al., 2017; Driedonks et al., 2018). However, despite the efforts, no major breakthrough has been achieved for imparting heat resistance in tomato.

In recent years, the rapid advances in genotyping and its decreasing cost offer widespread applications in breeding programs. In this study, high-throughput genotyping by the means of GBS approach, was performed to implement a new breeding selection scheme following a "Genomic selection" approach improving genetic gain for heat tolerance in tomato. The genetic architecture of heat tolerant traits was also investigated by QTL analysis followed by a gene variant characterization to identify potential candidates involved in heat tolerance.

### 3.4.1 GS model construction and validation

Our GS models, based on the genotyping and phenotyping of a F4 segregating population under heat stress, were able to predict the performance of not phenotyped F5 individuals further validated in F6 offspring. Indeed, although tomato pioneer GS studies have been conducted for yield-related traits (Duangjit et al., 2016; Yamamoto et al., 2016) also improving soluble solid content and total fruit weight (Yamamoto et al., 2017), there are no evidence on the analysis of these traits in heat stress conditions. Recent studies have demonstrated that the establishment of GS experiment optimal parameters is crucial for a reliable prediction (Robertsen et al., 2019). Selection response mainly depends on the TRN size, relatedness of the TRS and TST, marker density, precision of the phenotyping. In our GS experiment the optimal TRN was characterized by 90 and 100 (for YP and SSC respectively) F4 individuals.

Several works in literature explored the effect of the TRN size (Goddard et al., 2009; Sarinelli et al., 2019) showing that highest levels of accuracy have been reached with large TRNs (Meuwissen et al. 2013; Sarinelli et al., 2019). However, the optimal TRN size seems to be highly influenced by the relatedness of the TRS and TST (Schoop et al., 2017; Cericola et al., 2017; Edwards et al., 2019). In our work, also with relatively small populations, good predictions were obtained. The accuracy of our optimized GS models for the two traits of interest was 0.729 for YP and 0.715 for SSC. Our accuracies were higher than values (YP=0.3 and SSC=0.6) obtained, with a comparable TRN size (96 individuals), by Yamamoto et al. 2017. However, in our experiment TRNs were composed by full-

sib individuals related to TRS. The accuracy of a GS model when the TRS is fully related to the TST is normally between 0.4 and 0.7 (Bassi et al., 2016). TRS unrelated or distantly related to the TST, tends to give less accurate prediction (between 0.2 and 0.4). This is consistent with data obtained by Yamamoto et al. (2017) that employed a TRN population composed by 96 tomato unrelated varieties. Indeed, when the TRS and TST are unrelated, marker effects could be inconsistent due to the presence of different alleles, allele frequencies, and linkage phases.

The goal of GS model construction is to capture as much of the genetic variation as possible, and it was widely reported that a higher marker density will improve prediction accuracy (Zhang et al., 2019). Rapid advances in sequencing technologies are providing highly informative markers at very low cost, making sequence-based genotyping a very attractive approach to rapidly characterize genomes and populations. In this regard, Genotyping-by-sequencing (GBS) used in this study, have generated high-density markers efficiently and inexpensively, although with moderate rates of missing data. In our work a study of the effects of Training Set (TRS) size and GBS marker subsets filtering (by MAF and PEMV) on the model accuracy was performed. Interestingly, the most accurate prediction was obtained with the smallest marker subset in terms of marker density. This is an important point to take in account, since the use of fewer markers would result in significant cost reductions for genotyping, but would also impact the extent of LD that is picked up by the prediction models.

In previous studies, Yamamoto and collaborators (2017) used a set of 96 big-fruited F1 tomato varieties to develop GS models, and the segregating populations obtained from crosses were used to validate the models. Consequently, the GS models were used to predict parental combinations generating superior hybrids based on the values of approximatively 20 expected progeny phenotypes for soluble solids content and total fruit weight. The phenotypic diversity of evaluated progenies was narrow and contrasting indications were obtained for future parental crossing combinations. In our work, GS was implemented in a Single Seed Descendent (SSD) scheme where each generation derive from the former, taking only one seed from each parent plant. This scheme takes advantage of the gain increase generated by shortening the cycle length and offering great benefits especially in situations where simultaneous selection for several characteristics with different heritability is required (Kanbar et al., 2011). In the classical SSD scheme, no selection is conducted until the last generation (generally F6-F7), so the phenotyping of a larger number of lines could be challenging. The integration of the GS approach in the SSD scheme resulted in reducing the number of auto-selfing generations shortening the last two. Because the prediction accuracy is generally higher when LD (linkage disequilibrium) is high, an increase of the breeding gains was expected, applying GS in the earliest heterozygous segregating generations (i.e., F2-F4). Therefore, F4 generation was used for

developing the GS model, and subsequently GS prediction to assist selection in the next generations. Genomic data accurately track the best performing plants along the generations and the approach successfully lead to the selection of individuals with the highest GEBV in F5 and F6.

In previous studies potential assessment of the GS in tomato breeding was performed based on cross-validation using only training data (Duangjit et al., 2016; Yamamoto et al., 2016) or without cross-validation (Liabeuf et al., 2018). Our model was validated in F6 generation, confirming that our approach is promising since the correlation value between the GEBVs of F5 predicted individuals and the real values of F6 phenotyped individuals is in agreement. Our study displayed that GS could reduce the expenditures required for tomato breeding. Although more studies are needed to test GS in actual breeding programs, our results highlight GS as a promising strategy for future tomato breeding.

### 3.4.2 QTL analysis revealed candidate genes putatively involved in heat stress tolerance

In order to characterize the F4 population previously characterized at phenotypic level (see chapter II) also at genotypic level, further analyses were performed on the genomic data obtained. We found a high density of SNPs on chromosomes 4, 5, 6 and 11. To better investigate the genomic relatedness of individuals based on whole genome SNP information and to detect specific meaningful SNP regions on these chromosomes, a circos plot analysis was performed. This analysis allowed us to discover two extremely interesting regions on chromosome 4 and chromosome 6 associated with traits of interest (YP and SSC).

Subsequently, the information gathered by the GS model were further investigated in order to perform a QTL analysis based only on the more significant SNPs. Indeed, among all SNPs used for the model construction only those with higher SNP effect calculated by the model were used. In this way we used only the meaningful SNPs, drastically reducing the number of SNPs for the analysis. Conventional mapping allowed to identify several tomato QTLs associated with heat tolerance (Lin et al., 2006; Grilli et al., 2007; Xu et al., 2017; Driedonks et al., 2018; Wen et al., 2019) and these studies provided the basis for understanding the genetic architecture of heat tolerance traits. Our combined approach allowed to detected 5 significant QTLs for YP under heat stress, on chromosome 5, 6, 8, 9 and 11 (qYP1, qYP2, qYP3, qYP4, qYP5) and 1 significant QTL for SSC on chromosome 6 (qSSC1). Xu et al. (2017) and Driedonks et al. (2018) also detected a heat-tolerance QTL on chromosomes 11 related to pollen viability trait and a QTL on chromosome 8 related to inflorescences number. Besides, on chromosome 5, Wen et al. (2019) and Driedonks et al. (2018) found 2 QTLs related to a physiological traits and style protusion respectively and a QTL on chromosome 9 related to heat-tolerant anther length trait was reported from the former authors (Driedonks et al., 2018).

However, no heat-tolerance QTLs on chromosomes 6 were reported although our analysis revealed the highest LOD Scores for the two analyzed traits on this chromosome. Furthermore, our findings revealed that, the QTLs on chromosome 6 for the two analyzed traits mapped approximately to the same genomic region. This co-localization perfectly mirrored the negative correlation found between the phenotypic traits analyzed themselves (chapter II). Indeed, when the YP and SSC phenotypes were plotted against the genotypes, opposite homozygous states were observed.

The heat-tolerance QTLs on chromosome 6 covered a large area with a large number of candidate genes. In order to identify genes putatively involved in YP and SSC under heat stress, a variant annotation was performed to associate each variant of identified QTL to the annotated genes and to predict their effect on the protein function using a strategy similar to the one applied by Driedonks et al. (2018) and Wen et al. (2019). Among all the identified SNPs in the QTL region for YP trait, those that delimited the ends of interval (V66858 and V84318) resulted quite interesting since they are located on upstream region of the R2R3 trascription factor myb 59 (SlMYB59) and the SKP1-like protein respectively. Although the MYB59 gene is not yet functionally characterized in tomato, an orthologous to MYB59 gene was found in Arabidopsis (AtMYB59) (https://solgenomics.net/). AtMYB59 is involved in the development of secondary cell walls and vasculature in *A. thaliana* stems (Oh et al., 2003), in jasmonate signalling (Hickman et al., 2017) and in the response to K deficiency (Nishida et al., 2017), as well as to regulate growth and the cell cycle in roots (Mu et al., 2009). More recently, Du et al. and Fasani et al. (2019) highlighted the key role of AtMYB59 in multiple stress response since several genes involved in abiotic and biotic stress are modulated in the *myb59* mutant already in control conditions. On the other hand, little is known about the skp1-like protein. A blast analysis allowed us to identify as orthologous of the identified skp1-like protein the ASK2 in Arabidopsis. The Arabidopsis genome contains 21 SKP1-like genes called ASK (for Arabidopsis SKP1-like), among which only ASK1 and ASK2 has been characterized in detail (Fuquan Liu, 2004). It has been demonstrated that ASK1 and ASK2 play vital roles in embryo-genesis and postembryonic development, since *ask1ask2* double mutant showed substantial delay in embryogenesis and lethality in seedling growth. However, further analysis needed to confirm the role of these two proteins in tomato.

The candidate genes detected in the present study can be functionally exploited by TILLING or CRISPR-Cas9 approaches. If the candidate genes will confirm to be involved in the heat-tolerance response it could be then transferred to cultivated tomatoes to improve performances under high temperatures.

# CHAPTER IV

# EXPLORING SlDOF TRANSCRIPTION FACTORS FAMILY TO IDENTIFY POTENTIAL REGULATORS OF HEAT STRESS TOLERANCE IN TOMATO

## 4.1 INTRODUCTION

Tomato is a horticultural crop of major economic importance worldwide though the continuous manifestations of biotic and abiotic stresses during the different growth stages led severe yield reductions (Mittler, 2006; Prasad et al., 2011). In future, the temperate regions are expected to be heavily impacted by climate change (Cramer et al., 2018). In particular, in semiarid temperate regions, the co-occurrence of heat stress with drought can seriously limit tomato cultivation.

Important advancements have been achieved in our understanding of transcriptional regulation, signal transduction, and gene expression in plant responses to abiotic stresses (Zhu et al., 2010; Zafar et al., 2019). Indeed, the identification of regulatory genes involved in the abiotic stress tolerance will provide new molecular targets to implement breeding programs. Transcription factors (TFs) and co-transcriptional regulators acting in various signaling cascades can adjust the molecular response to environmental cue. Various TF members belonging to the MYB, WRKY, NAC, DOF, AREB/ABF (ABA response-element binding factor), GBFs (G-box binding factors), and AP2/ERF families (Bostock et al., 2014) have been characterized for their involvement in abiotic stress response (table 1). In rice, for example, overexpression of a NAC encoding gene, SNAC1, promotes yield increasing under drought stress condition (Hu et al., 2006). Similarly, functional analysis of the PbeNAC1 in *Pyrus betulifolia*, revealed that this gene is involved in the regulation of cold and drought stress tolerance (Jin et al., 2017). Overexpression of OsMYB55, a rice MYB encoding gene in transgenic maize resulted in improved plant growth as well as decreased the negative effects of drought and high temperature (Casaretto et al., 2016). Wei et al. (2017) demonstrated that CiMYB3 and CiMYB5 cloned from *Cichorium intybus* were involved in various abiotic stresses response. WRKY6 acts as a negative regulator in both transgenic Arabidopsis and Cotton during drought and salt stress (Li et al.,

2019). SlWRKY39 confer both drought and salt tolerance in tomato by activating abiotic stress genes (Sun et al., 2015). More recently, Dof factors emerged as mayor stress response players since they are involved in several biotic and abiotic stress response (Corrales et al., 2014; Corrales et al., 2017; Ewas et al., 2017).

| Trascription factor | Transformed species | Stress tolerance | Reference |
|---|---|---|---|
| OsABF3 | *Triticum aestivum* | Drought, Salt | Lu et al., 2009 |
| PtrABF | *Nicotiana nudicaulis* | Drought, osmosis | Huang et al., 2010 |
| GmWRKY54 | *Glicine max* | Salt, drought | Wei et al 2019 |
| GmWRKY27 | *Glicine max* | Drought, salt | Wang et al., 2015 |
| AtCBF3(AP2/ERF) | *Oryza sativa* | Drought, salt, cold | Oh et al 2005 |
| AtNAC042 | *Arabidopsis thaliana* | Drought, heat | Shahnejat et al 2012 |
| SNAC1 | *Oryza sativa* | Increased yield under drought stress | Hu et al., 2006 |
| PbeNAC1 | *Pyrus betulifolia* | Cold, drought | Jin et al.,2017 |
| AtDREB2C | *Arabidopsis thaliana* | Salt, cold | Lee et al, 2010 |
| OsDREB2B | *Arabidopsis thaliana* | Drought, heat | Chen et al, 2008 |
| OsMYB4 | *Arabidopsis thaliana, Solanum lycopersicum* | Drought, cold | Vannini et al 2007.Pasquali et al., 2008. Agarwal and Jha 2010 |
| OsMYB55 | *Zea Mais* | Heat, drought | Casaretto et al., 2016 |
| CiMYB3, CiMYB5 | *Cichorium intybus* | Abiotic stress response | Wei et al. (2017) |
| GhWRKY6 | *Arabidopsis thaliana, Cotton* | Drought, Salt | Li et al., 2019 |
| SlWRKY39 | *Solanum lycopersicum* | Drought, Salt | Sun et al., 2015 |
| CDF3 (Dof) | *Arabidopsis thaliana* | Drought, Cold | Corrales et al., 2017 |
| SlCDF1, SlCDF3 | *Solanum lycopersicum* | Drought, Salt | Corrales et al., 2014 |
| SlTDDF1 | *Solanum lycopersicum* | Drought, Salt | Ewas et al., 2017 |

**Table 1. Transcription factors involved in regulating plants in response to multiple abiotic stresses.**

The Dof (DNA-binding with one finger) proteins constitute a plant-specific family of TFs harboring a DNA-binding domain, which forms a single zinc-finger (Noguero et al., 2013). These proteins are characterized by an highly conserved DNA-binding domain at the N-terminal composed of 52 amino acid residues structured as a Cys2/Cys2 (C2/C2) zinc finger that recognizes a cis-regulatory element containing the common core sequence 5′-(T/A)AAAG-3′ (Yanagisawa, 2004). However, these proteins also contain a more variable C-terminal transcriptional regulation domain containing diverse

amino acid sequences (Yanagisawa, 2001; Gupta et al., 2015). The N- and C-terminal regions of the Dof proteins contribute to their bi-functional roles in DNA binding and protein-protein interactions to regulate the expression levels of the target (Gupta et al., 2015). In addition to the highly conserved Dof domain, up to 25 conserved domains have been identified in this gene family. These additional domains result in a high divergence in the structure of the genes between the different groups or subgroups (Cai et al., 2013).

The Dof gene family size widely vary in the genomes of plant species. It ranges from 9 genes in *Physcomitrella patens* to 54 Dof genes in maize (Gupta et al., 2015). The first identified Dof gene, ZmDof1 was found to play a role in light-regulated gene expression and it affects light response and nitrogen assimilation (Yanagisawa and Izui, 1993; Yanagisawa and Sheen, 1998). Subsequently, a large number of Dof genes were reported to be involved in a variety of plant-specific biological processes, such as seed germination (Santopolo et al., 2015), pollen development (Peng et al., 2017), endosperm development (Wu et al., 2019), fruit ripening (Feng et al., 2016), flowering time control (Wu et al., 2017), plant architecture (Wu et al., 2015), carbon and nitrogen metabolism (Santos et al., 2012), and responses to plant hormones (Rymen et al., 2017; Lorrai et al., 2018; Qin et al., 2019), as well as various abiotic and biotic stress responses (Su et al., 2017; Zang et al., 2017). In this regard, for example, overexpression of Arabidopsis CDF3 showed its involvement in both flowering time control and abiotic stress tolerance (Corrales et al., 2017). In tomato, the overexpression of TDDF1, a Dof gene inducing early flowering displayed higher resistance to drought, salt, and late blight caused by *Phytophthora infestans* (Ewas et al., 2017).

In rice, overexpression of OsDof15 reduced the sensitivity of roots to salt stress via restricting ethylene biosynthesis, suggesting that OsDof15-mediated ethylene biosynthesis plays a role in the inhibition of primary root elongation by salt stress (Qin et al., 2019). These findings demonstrate that the Dof proteins are playing important roles in the plant growth and development, but also in several biotic and abiotic stress response.

In tomato, 34 Dof proteins, classified in 4 classes and distributed in 11 chromosomes have been identified (Cai et al., 2013, table 2). However, despite the importance of this gene family during plant growth, only a small number of members have been functionally characterized in tomato. A group of five tomato Dof genes, homologous to Arabidopsis Cycling Dof Factors (CDFs) are reported to be mainly involved in the control of flowering time (Corrales et al., 2014). Among these, overexpression of the cycling Dof factor (CDF) TF designated as CDF3 resulted also in increased tomato biomass production and higher fruit yield under salt stress (Renau-Morata et al., 2017).

In 2017, Ewas et al., have characterized the TDDF1 gene showing that this Dof was involved in circadian regulation and salt and drought stress resistance. More recently, Rojas Gracia et al. (2019)

have shown the role of SlDof10 gene in vascular tissue elongation specifically during reproductive development. Therefore, additional work is required to fully understand the role of Dof genes during tomato plant growth and development and in response to the abiotic stress.

| Unigene name† | Locus gene | Predicted gene structure‡ | Chomosome and location | | Group§ | Size (AA) | pI | Mw (kD) |
|---|---|---|---|---|---|---|---|---|
| SGN-U577255 | Solyc01g096120 | __▼_Dof_▼_ | 1 | 78996952..78998756 | Class 1 | 276 | 9.32 | 29.8 |
| SGN-U277346 | Solyc02g065290 | ___Dof___ | 2 | 31027228..31026617 | Class 3 | 203 | 9.02 | 21.9 |
| SGN-U567927 | Solyc02g067230 | _▼_Dof__ | 2 | 31993302..31990580 | Class 4 | 467 | 4.83 | 51.3 |
| SGN-U599962 | Solyc02g076850 | ___Dof___ | 2 | 36554424..36554915 | Class 4 | 163 | 7.6 | 18.2 |
| SGN-U575537 | Solyc02g077950 | ___Dof___ | 2 | 37348924..37350063 | Class 3 | 379 | 8.16 | 41.6 |
| SGN-U575558 | Solyc02g077960 | ___Dof___ | 2 | 37366575..37365676 | Class 2 | 299 | 9.63 | 32.2 |
| SGN-U595859 | Solyc02g078620 | ___Dof___ | 2 | 37850975..37851829 | Class 3 | 284 | 5.98 | 32.3 |
| | Solyc02g088070 | _▼_Dof__ | 2 | 44845428..44848303 | Class 4 | 462 | 5.98 | 51.2 |
| | Solyc02g090220 | ___Dof___ | 2 | 46424050..46424754 | Class 3 | 234 | 8.71 | 24.8 |
| SGN-U584226 | Solyc02g090310 | ___Dof___ | 2 | 46468203..46467421 | Class 2 | 260 | 8.83 | 28.6 |
| SGN-U571096 | Solyc03g082840 | _▼_Dof___ | 3 | 46233684..46232681 | Class 1 | 258 | 8.55 | 28.1 |
| SGN-U568164 | Solyc03g112930 | _▼_Dof___ | 3 | 57277135..57275421 | Class 1 | 333 | 6.38 | 36.7 |
| SGN-U578970 | Solyc03g115940 | _▼_Dof___ | 3 | 59569502..59567362 | Class 4 | 501 | 5.35 | 54.5 |
| SGN-U568282 | Solyc03g121400 | ___Dof___ | 3 | 63560019..63559387 | Class 2 | 210 | 9.20 | 22.6 |
| SL2.30sc04135 | Solyc04g070960 | ___Dof___ | 4 | 55346815..55346012 | Class 1 | 267 | 4.79 | 30.4 |
| SL2.30sc04135 | Solyc04g079570 | ___Dof___ | 4 | 61583624..61582905 | Class 3 | 239 | 6.42 | 24.4 |
| SGN-U570634 | Solyc05g007880 | _▼_Dof___ | 5 | 2322103..2319855 | Class 4 | 441 | 6.13 | 49.1 |
| | Solyc05g054510 | _▼_Dof___ | 5 | 63567003..63565916 | Class 3 | 298 | 9.46 | 33.7 |
| SGN-U597361 | Solyc06g005130 | _▼_Dof___ | 6 | 158970..157503 | Class 1 | 306 | 6.59 | 34.4 |
| SGN-U583873 | Solyc06g062520 | ___Dof___ | 6 | 35829346..35828768 | Class 2 | 192 | 8.47 | 21.0 |
| SGN-U582777 | | ___Dof___ | 6 | 39537684..39538604 | Class 3 | 306 | 5.14 | 33.9 |
| SGN-U577422 | Solyc06g069760 | _▼_Dof___ | 6 | 39835980..39833891 | Class 4 | 492 | 6.12 | 53.5 |
| | Solyc06g071480 | ___Dof___ | 6 | 40385794..40387296 | Class 1 | 268 | 8.00 | 30.1 |
| SGN-U563536 | Solyc06g075370 | _▼_Dof___ | 6 | 43178563..43176963 | Class 1 | 303 | 7.12 | 33.7 |
| SGN-U574638 | Solyc06g076030 | _▼_Dof___ | 6 | 43593315..43592150 | Class 3 | 334 | 9.26 | 36.0 |
| SGN-U590771 | Solyc08g008500 | ___Dof___ | 8 | 2894917..2895989 | Class 1 | 227 | 8.11 | 25.5 |
| SGN-U591256 | Solyc08g082910 | _▼__Dof_▼_ | 8 | 62732144..62733841 | Class 1 | 318 | 7.03 | 35.2 |
| SGN-U601651 | Solyc09g010680 | _▼__Dof_▼_ | 9 | 4000596..4001738 | Class 3 | 330 | 9.42 | 35.4 |
| SGN-U579360 | Solyc10g009360 | ___Dof_▼_ | 10 | 3425751..3424646 | Class 1 | 261 | 9.33 | 28.5 |
| | Solyc10g086440 | _▼_Dof___ | 10 | :64586916..64585638 | Class 3 | 362 | 9.41 | 38.9 |
| SGN-U600655 | Solyc11g010940 | ___Dof_▼_▼_ | 11 | 3989496..3987675 | Class 2 | 374 | 6.54 | 40.8 |
| SGN-U571346 | Solyc11g066050 | ___Dof___ | 11 | 48809837..48809043 | Class 3 | 264 | 9.40 | 28.7 |
| SGN-U573048 | Solyc11g072500 | _▼_Dof___ | 11 | 52780331..52781968 | Class 1 | 282 | 7.09 | 32.1 |
| | Solyc00g024680 | _▼_Dof___ | 0¶ | 12224977..12225584 | Class 3 | 173 | 9.63 | 18.6 |

**Table 2. Annotation of predicted SlDof genes in tomato. Relative position of introns ( ▼ ) with respect to the Dof domain. AA, amino acids; pI, isoelectric point of the deduced polypeptide; Mw, molecular weight (by Cai et al., 2013).**

This species is both important as food crop than as model plant species that has been used extensively for studying gene functions. Its genetic features combined with readily available technologies (site specific cleveage enzymes, genome sequences, transformation methodology) makes tomato an ideal candidate for gene editing. Genome engineering techniques for dissecting out the molecular function of TFs have been dramatically improved by the genome editing technology. In particular, clustered regulatory interspaced short palindromic repeats “CRISPR”-Cas9 system resulted as the most effective system used in editing plant genomes (Cong et al., 2013). In the

CRISPR/Cas9 system, Cas9 functions as a Cas9-gRNA ribonucleoprotein complex and uses its DNA endonuclease activity to induce the cleavage of the genome, which is targeted by gRNA. Double-strand breaks sometimes induce insertions and deletions at the target site, leading to frameshift mutations of genes. The simplicity of vector construction and the versatility of experiments design has dramatically increased the number of genome editing studies in plants including tomato (Klap et al., 2017; Swinnen et al., 2020).

In this work, with the aim to identify regulatory tomato SlDof genes involved in the abiotic stress tolerance, a genome scan for Dof protein domains and a phylogenetic analysis of this gene family including Arabidopsis members was performed. Among these, the Dof factor STOMATAL CARPENTER 1 (SCAP1/Dof5.7) resulted of interest since it regulates essential processes of stomatal guard cell (GC) maturation and functions as a key transcription factor regulating the final stages of guard cell differentiation (figure 1). SCAP1-mutants displayed altered levels of transcripts of multiple genes directly involved in stomatal functioning and morphogenesis and furthermore are defective in some mechanical properties of the GC cell wall (Negi et al., 2013; Castorina et al., 2016). These findings highlight the key role of SCAP1/Dof5.7 in abiotic stress response. Once the orthologous of SCAP1 has been identified in tomato, targets suitable for a genome editing approach were designed in order to functionally characterize this Dof factor. In addition, a CRISPR/Cas9 experiment was carried out using the Golden Braid modular system.

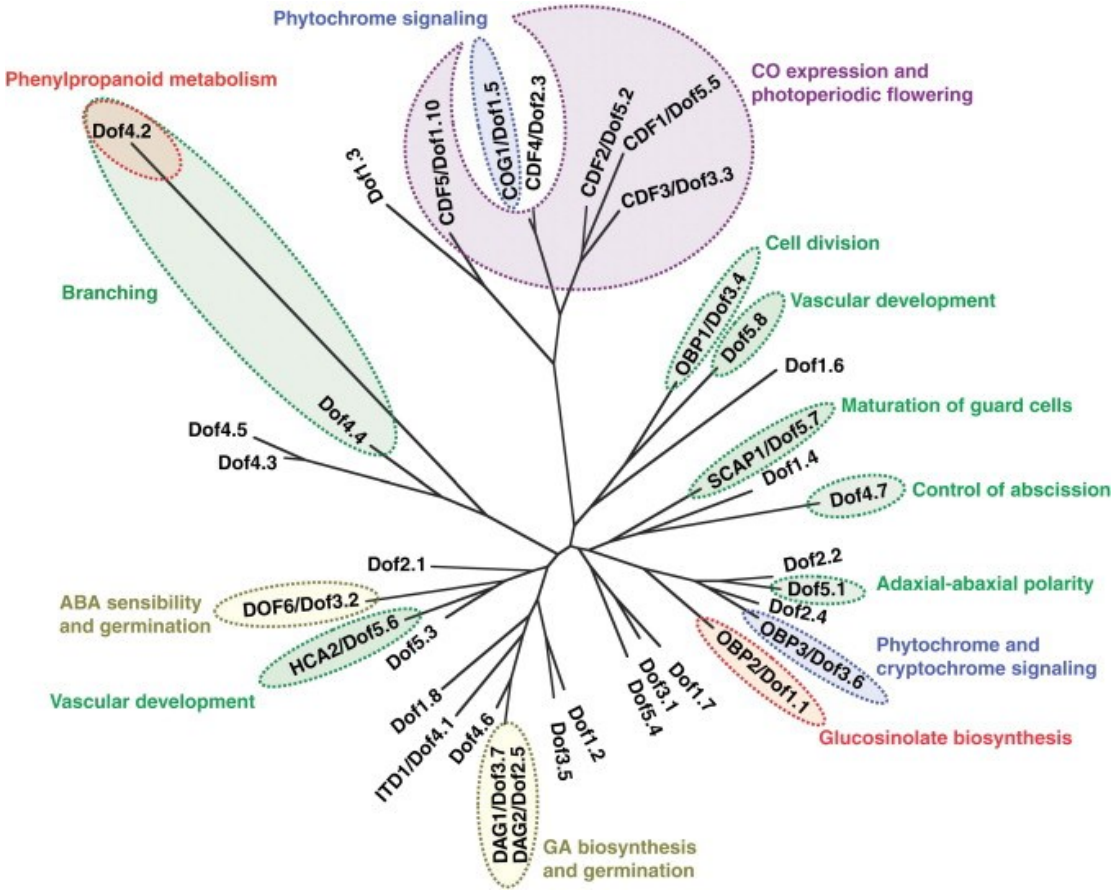



**Figure 1. Physiological functions of the Arabidopsis Dof transcription factors. (by Yanagisawa et al., 2015).**

## 4.2 MATERIALS AND METHODS

### 4.2.1 Annotation of DOF-encoding genes

To identify DOF encoding genes, we scanned the *Solanum lycopersicum* (SL3.0) proteome with Hidden Markov Model (HMM) of zf-Dof (Pfam: PF02701) using HMMER v3.0 with default parameters (Finn et al., 2011). The seed sequence of the zf-Dof domain was retrieved from Pfam v31.0 (http://pfam.xfam.org). The protein domain architecture of HMMER outputs was annotated using IntetProScan v5 (Jones et al., 2014).

### 4.2.2 Phylogenetic analysis of DOF-encoding genes

For comparative propose, the 34 *Solanum lycopersicum* full-length protein sequences identified and 36 *Arabidopsis thaliana* Dof proteins obtained from the TAIR database (https://www.arabidopsis.org) were aligned with ClustalW (Larkin et al., 2007), using default setting. Evolutionary analyses were conducted using MEGA7 (Kumar and Tamura, 2015). The phylogenetic relationships of Dof-encoding genes were inferred using the maximum likelihood method based on the Jones et al. (1992) mode. Model with the lowest BIC (Bayesian Information Criterion) score was considered to describe the substitution pattern. The bootstrap consensus tree inferred from 100 replicates was taken to represent the evolutionary history of the sequences analysed. The trees were drawn to scale, with branch lengths measured in the number of substitutions per site. Finally, to reroot the phylogenetic trees, *Triticum aestivum* DOF4 (accession number FJ687390) was used.

### 4.2.3 Targets selection and vector construction

#### 4.2.3.1 gRNA target sequences design

Benchling was used for the gRNAs design. Parameters for identification of possible off-targets were: less than five mismatches (L0 = 4) or less than four mismatches (L1 = 3), if one mismatch is found in seed sequence, gRNA with PAM ‘NGG and start with ‘G’, on-target score and off-target score >45%. Once all possible gRNA sequences in target gene were found, the protein sequences of the top-ranked targets (https://solgenomics.net) was loaded in prosite (https://prosite.expasy.org/) to identify the functional domain (Zinc-finger-DOF-type). Only the gRNAs able to target the conserved functional domain were chosen. In the end, to validate the selected gRNAs, Cas-OFFinder tool (http://www.rgenome.net/cas-offinder/) was used to search throughout the entire genome of interest for possible off-target sequences that differ by up to 2 nucleotides from the on-target sequences.

Parameters used were: maximum number of mismatches: 3; off-target sites with 2 bp deletions or insertions (DNA/RNA bulges).

#### 4.2.3.2 Strains and Growth Conditions

*Escherichia coli* DH5α was used for cloning the sequences and *Agrobacterium tumefaciens* strain LBA4404 was used for plant transformation experiments. Both strains were grown in Luria-Bertani medium under agitation (200 rpm) at 37°C and 28°C, respectively. Kanamycin (50 µg mL−1), ampicillin (100 µg mL−1) and spectinomycin (50 µg mL−1) were used for *E. coli* selection. Rifampicin, streptomycin, and kanamycin were also used for *A. tumefaciens* selection at 50 µg mL−1. 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside acid (40 µg mL−1) and isopropylthio-β-galactoside (0.5 mM) were used on Luria-Bertani agar plates for the white/blue selection of clones.

#### 4.2.3.3 Restriction-Ligation Assembly Reactions

Restriction-ligation reactions were set up as described elsewhere (Sarrion-Perdigones et al., 2011) using *BsaI, BsmBI*, as restriction enzymes (Thermo Fisher) and T4 Ligase (Thermo Fisher). Reactions were set up in 25 cycle digestion/ligation reactions (2 min at 37°C, 5 min at 16°C). One microliter of the reaction was transformed into *E. coli* DH5α electrocompetent cells, and positive clones were selected in solid medium. Plasmid DNA was extracted using the E.Z.N.A. Plasmid Mini Kit I (Omega Bio-Tek). Assemblies were confirmed by restriction analysis and sequencing. All plasmid digestions were compared with in silico predicted plasmid digestions on Benchling (https://benchling.com).

#### 4.2.3.4 GB segments construction

The GoldenBraid 2.0 Kit, including the complete set of vectors required to perform GoldenBraid cloning was used (https://www.addgene.org/kits/orzaez-goldenbraid2/). All vectors used in this work have been designed using Golden Braid 3.0 system following the described assembly strategy (Sarrion-Perdigones et al., 2013; Sarrion-Perdigones et al., 2014; Vazquez-Vilar et al., 2015; Vazquez-Vilar et al., 2016). Sequences of GB-Parts are accessible at GB cloning website (https://gbcloning.upv.es/) using the GB database ID.

#### 4.2.3.5 gRNAs assembly in α-Level Plasmids (Multipartite Reaction)

The optimal primer combinations for the gRNAs selected were analysed on Benchling.

The 20nt sgRNA39 and sgRNA558 were assembled as follows: sgRNA39_F, sgRNA39_R and sgRNA558_F, sgRNA558_R (supplementary table 2), were resuspended in water to final concentrations of 1 µM and incubated at room temperature for 30 min. The BsaI restriction–ligation

reactions were set up in 10 µl with 0.6 µl of primers mix, 75 ng of GB1001 (U626 promoter), 75 ng of GB0645 (scaffold RNA), 3u BsaI, 3u of T4 ligase, 1 µl of ligase buffer and 75 ng of pDGB3_alfa1 destination vector for sgRNA39 and pDGB3_alfa2 for sgRNA558. After 25 cycles × (37 °C 2 min, 16 °C 5 min) one microliter of the reaction was transformed into *E.coli* electrocompetent cells and the number of white colonies growing on agar plates counted. All gRNA constructs were validated by RE-analysis, analyzed by sequencing and confirmed correct. Plasmid digestions were compared with in silico predicted plasmid digestions on Benchling.

#### 4.2.3.6 Multigene Assembly in Ω Level Plasmids (Binary Reaction)

GB parts assembled in α-level plasmids were combined in Ω-level plasmids. The assembled reaction contained 75 ng of the Ω-level destination vectors (pDGB3_omega1, pDGB3_omega2), 75 ng of the TUs to be assembled, 3u BsmBI and 3u T4 DNA ligase in a final volume of 10 µl. After 25 cycles × (37 °C 2 min, 16 °C 5 min), one microliter of the reaction was transformed into *E.coli* electrocompetent cells and the number of white colonies growing on agar plates counted. All constructs were validated by RE-analysis and sequencing. Plasmid digestions were compared with in *silico* predicted plasmid digestions.

#### 4.2.3.7 Multigene Assembly in α-Level Plasmids (Binary Reaction)

In the end, composite parts GB assembled in Ω-level plasmids were combined in α-level plasmids in order to obtain the final constructs. The assembled reaction contained 75 ng of the α-destination vector, 75 ng of the TUs to be assembled, 3u BsaI, and 3u T4 Ligase in a final volume of 10 µL. After 25 cycles × (37 °C 2 min, 16 °C 5 min), one microliter of the reaction was transformed into *E.coli* electrocompetent cells and the number of white colonies growing on agar plates counted. All constructs were validated by RE-analysis and sequencing. Plasmid digestions were compared with *in silico* predicted plasmid digestions.

### 4.2.4 Plant Material and Growth Conditions

Seeds of *Solanum lycopersicum* (cv *Red setter*) were surface sterilized with 70% (v/v) ethanol for 1 min and then in 2% (v/v) sodium hypochlorite solution for 10 min, thoroughly washed, stratified for 48 h at 4 °C in the dark and sown on Murashige & Skoog solid medium pH 5.8 containing 30 g l−1 sucrose and 9 g l−1 agar. The plants were grown at 23 °C, under long-day photoperiod (16 h light, 8 h dark) in presence of cool white fluorescent light (110 µmol m−2 s−1).

### 4.2.5 *Agrobacterium tumefaciens*-mediated Transformation

Multigene constructs were transferred into *Agrobacterium tumefaciens* adding 15 ng of plasmids to 50 µL of *A.tumefaciens* electrocompetent cells and 500 µL of SOC medium Reactions were incubated in a 15 mL tube at 28 °C for 2 h with agitation. Different amount of reaction (50, 100, 200 and 500µL) were spread in LB plates containing kanamycin, streptinomycin and rifampicin and then incubate for 48 h in a 28 °C growing chamber. The positive colonies were validated by RE-analysis. *A. tumefaciens*-mediated transformations of the tomato (*Solanum lycopersicum*) cv *Red Setter* were performed according to Van Eck et al. (2006). In brief, 1 ml starter culture of the Agrobacterium was inoculated into 50 ml of LB media containing 50 mg /l rifampicin and 50 mg/l kanamycin and grown at 28°C. The culture was centrifuged at 5000 g for 10 min. The bacterial pellet was re-suspended in fresh MS media to an OD600 of 0.6 and used for the plant transformation procedure. Leaf discs were inoculated for 20 min in the *A. tumefaciens* suspension. Inoculated leaf discs were blotted dry on sterile filter paper, placed on cocultivation media (4.46 g MS salts with vitamins, 1 µM BAP, 30 g/l sucrose, pH 5.8) and incubated at 22°C for 2 days in the dark. After two days the leaf discs were transferred to regeneration media (4.46 g MS salts with vitamins, 1 µM BAP, 30 g/l sucrose, and 400 mg/l cefotaxime, pH 5.8), and kept at 26°C in a 16 h/8 h day/night cycle under coolwhite fluorescent lights and sub-cultured to fresh media every two weeks. Shoots began to develop after 4 weeks. Each site of shoot development was considered a separate transgenic event. Ten putative transgenic shoots were transferred to rooting media (2.23 g/l MS salts with vitamins, 20 g/l sucrose, mg/l kanamycin and 400 mg/l cefotaxime, and pH 5.8) in small 100 ml culture vessels.

## 4.3 RESULTS AND DISCUSSIONS

### 4.3.1 Phylogenetic analysis of the tomato Dof transcription factors

The Dof multigene family showed to play important roles in many biological processes during development and in response to environmental stimulus (Peng et al., 2017). In recent years, several Dof family genome-wide surveys have been conducted. In physic nut (*Jatropha curcas*) has been identified 25 Dof genes (Wang et al., 2018; Zou and Zhang, 2019), in peach (*Prunus persica*) 25 members (Chen et al., 2017), in eggplant (*Solanum melongena*) 29 (Wei et al., 2018), 33 in pepper (*Capsicum annuum*) (Kang et al., 2016; Wu et al., 2016), 35 in potato (*Solanum tuberosum*) (Venkatesh & Park, 2015), 36 in cucumber (*Cucumis sativus*) (Wen et al., 2016), 45 in cassava (*Manihot esculenta*) (Zou et al., 2019), 45 in pear (*Pyrus bretschneideri*) (Liu et al., 2019), 46 in rubber tree (*Hevea brasiliensis*) (Zou and Yang, 2019), 36 Watermelon (Zhou et al., 2019). The number of Dof members found in tomato (34) is in line with the number retrieved in other close species, suggesting that the Dof multigene families was conserved within botanical families. In this work, with the aim to identify regulatory Dof genes putatively involved in the abiotic stress response in tomato, a fine gene annotation and a phylogenetic analyses including tomato and Arabidopsis members were performed in order to identify orthologous Dof genes. In particular, for this analysis the full-length sequence of 34 *Solanum lycopersicum* and 36 *Arabidopsis thaliana* Dof- protein were employed. Results showed that each tomato Dof gene has at least one homologous gene in Arabidopsis (figure 2), suggesting that Dof genes might play similar roles in tomato and Arabidopsis. Similar results were obtained in other species such as watermelon, eggplant, potato (Zhou et al., 2019; Wei et al., 2018; Venkatesh & Park, 2015) confirming our findings.

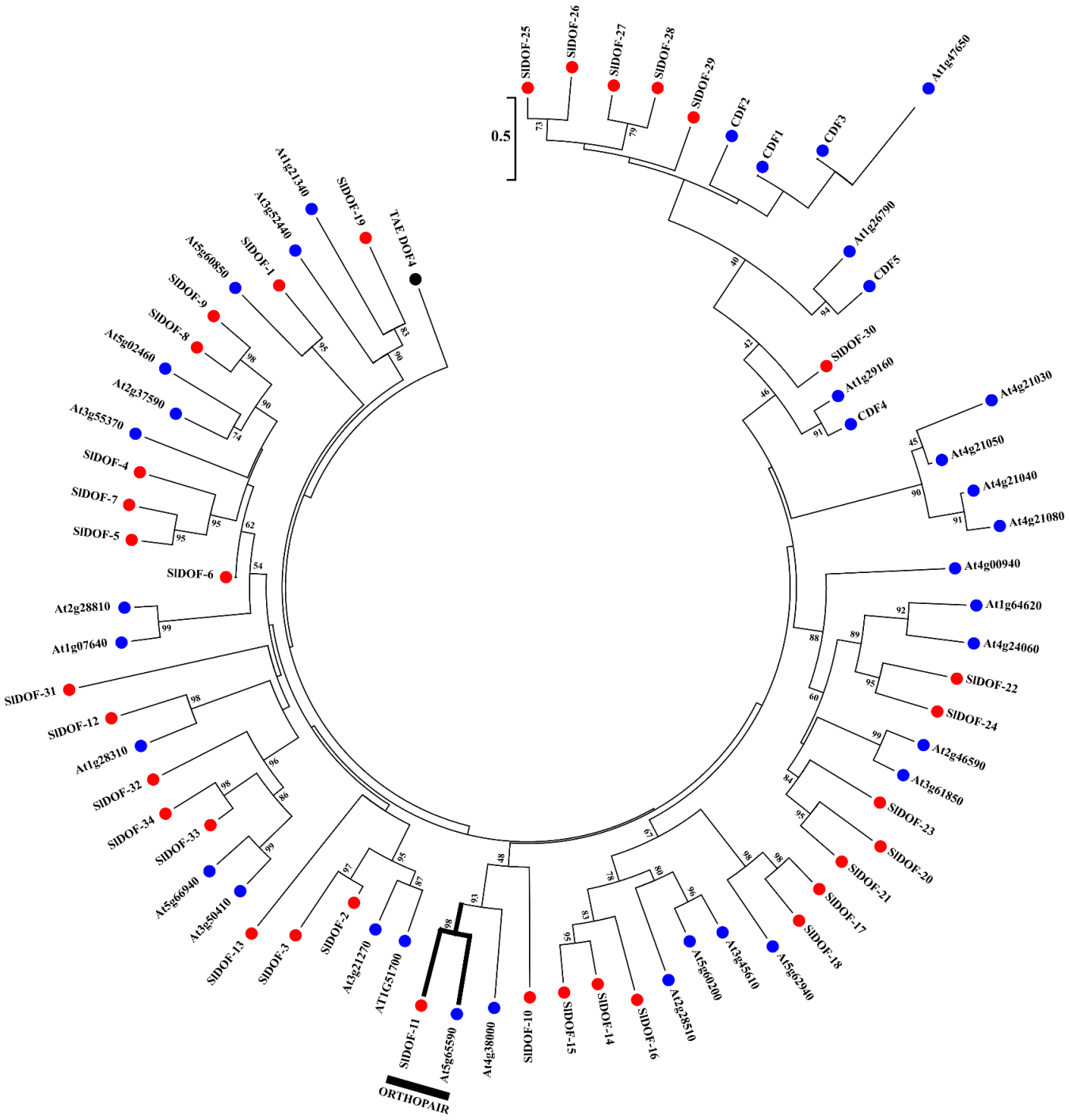



**Figure 2. Phylogenetic tree of Arabidopsis and tomato Dof genes. The full length amino acid sequences of AtDofs and SlDofs were aligned by ClustalW.**

In particular, 4 orthologous gene pairs between tomato and Arabidopsis were identified in our dataset. The number of the SlDofs and AtDofs pair-orthologous turns out to be lower, compared for example with the 10 eggplant - Arabidopsis orthologous base pairs (Wei et al., 2018) suggesting that a higher diversification occurred in the former species.

Among the 4 orthologous pairs, the SlDOF11 (Solyc11g090140.1) resulted the orthologous to the Dof factor STOMATAL CARPENTER 1 (SCAP1) in Arabidopsis. The structure of the SlDof11 is displayed in figure 3.

**SlDof11**


**Figure 3. Structure of the tomato SlDOF11**

### 4.3.2 SlDOF11 sequence investigation for designing suitable CRISPR/Cas9 sRNA guides

As a proof of concept, we decided to use the gene encoding SlDof11 as target in a *S.lycopersicum* gene-editing experiment, given that the loss of function in Arabidopsis SCAP1 impair stomatal opening and closing and repress the expression of genes involved in stomatal movement and morphogenesis (i.e., cell wall architecture). This transcription factor regulates the final stages of guard cell differentiation, indicating that SCAP1 regulates essential processes of stomatal guard cell (GC) maturation and functions. Therefore, SlDof11 could exert a key role in abiotic stress response.

A CRISPR/Cas9 based genome editing approach was conducted to characterize the SlDof11 function. The targeting of key traits related to abiotic stresses can be harnessed prolifically from CRISPR/Cas9 system. For example, OsRR22 and OsNAC041 have been edit in rice to increase salinity tolerance (Zhang et al., 2019). In corn the expression of ARGOS8 have improved through CRISPR/Cas based genome editing approach to enhance yield under drought stress drought tolerance (Shi et al., 2017). The tomato MAPK3 was edit through CRISPR/Cas9 confirming to be involved in adaptive abiotic stress responses (Duan et al., 2016; Wang et al., 2017; Huang et al., 2018). Recently, two abiotic-stress-responsive transcription facto r genes, encoding dehydration responsive element binding protein 2 (TaDREB2) and ethylene responsive factor 3 (TaERF3) were edit in wheat protoplasts by Kim et al. (2018). CRISPR/Cas9-induced knockdown of the tomato slagamous-like 6 (SlAGL6) gene responsible for the parthenocarpic phenotype, making tomato plants able to produce parthenocarpic fruits under heat stress (Klap et al., 2017).

The gRNAs design to target the DNA sequence of SlDof11 gene was a very critical point for producing edit plants since the specificity of the gRNAs can greatly influence the success of the experiment. We found more of 25 putative sgRNA sequences in the target gene. Suitable gRNA targets were selected taking into account target and off-target sites. Moreover, only gRNAs that started with 'G' were selected since 'G' at the transcription start site increase its transcription efficiency when a U6 promoter is employed in genome editing experiments. Following these rules, 9 putative gRNAs were selected for further analysis (table 3).

| Position | Strand | Sequence | PAM | On-Target score | Off-Target score |
|---|---|---|---|---|---|
| 39 | + | GGCGAAAGCCAGAGTTCTGG | TGG | 61.9 | 49.6 |
| 42 | + | GAAAGCCAGAGTTCTGGTGG | TGG | 57.3 | 48.4 |
| 112 | - | GTATTGGGAGAATCACATCT | AGG | 61.0 | 45.1 |
| 558 | + | GCTTTACGCAACGTTCCAAT | AGG | 49.5 | 48.8 |
| 1033 | + | GGATTTTTTTAACACGACAA | CGG | 61.8 | 88.1 |
| 1126 | + | GCTTTAGTGTCGTTTCTCTC | AGG | 45.3 | 48.2 |
| 1327 | + | GGATCAATGGGAGTTCATCA | TGG | 60.2 | 45.9 |
| 1335 | + | GGGAGTTCATCATGGTACAA | TGG | 51.3 | 47.6 |
| 1468 | - | GGTTTTGATCATCATGAAGA | GGG | 55.8 | 77.8 |

**Table 3. The 9 top-ranked gRNAs selected on Benchling. The position, strand, sequence, PAM and the percentage of On-Target and Off-Target score were reported.**

Guide RNAs targeting conserved functional domains can increase the likelihood that a mutation will compromise protein function (Shi et al., 2015; Giuliano et al., 2019). Therefore, only 4 target gRNAs located on the conserved Zinc-finger-Dof-type domain of SlDof11 protein sequence (gRNA in position 39, 42, 112 and 558) were further selected. In particular the gRNA39, gRNA42 and gRNA112 targeted Cas 9 to the exon 1 of the gene whereas the gRNA558 targeted the Cas9 to the exon 2. To better investigate the performance of such targets, we predicted the off-target sites with maximum mismactches of 2 basis and off-target sites with 2 bp deletions or insertions (i.e. DNA/RNA bulges).

The 2 gRNAs with lower off-target sites were: GGCGAAAGCCAGAGTTCTGG (sgRNA39) and GCTTTACGCAACGTTCCAAT (sgRNA558). The predicted off-target sites were reported in the supplementary table 1. These 2 gRNAs target the Cas9 to two different exons that encode functional protein domains (figure 4) enlarging the chance to ablate gene function than simply targeting a 5'exon (Shi et al., 2015, Giuliano et al., 2019).

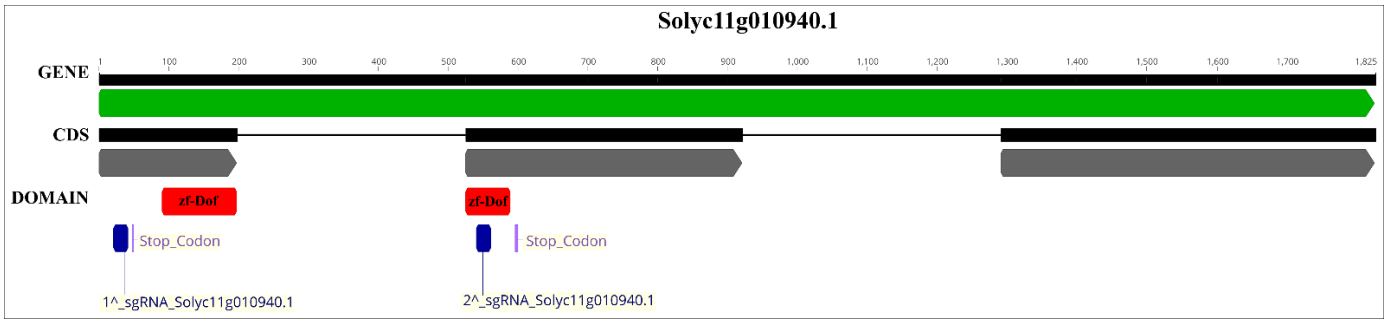



**Figure 4. The position of the 2 gRNAs selected were displayed. The 2 gRNAs (in blue) target the Cas9 to two different exons (in grey) that encode functional protein domain also located on the two different exons (in red).**

### 4.3.3 CRISPR/Cas9-Based Genome Editing of the Transcription Factor SlDOF11

Different editing strategies were checked to evaluate the effects of different mutations targeting the SlDOF11 gene. Single guide (sgRNA39 and sgRNA558) targeting unique regions and paired gRNA39 and gRNA558 constructs were tested. Indeed, although sgRNA design is based on a relatively simple 5′-N(20)-NGG-3′ targeting rule (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013), the efficiency of different gRNAs could vary in the cell. Targeted genomic deletions by CRISPR/Cas9 have been observed in numerous studies. Short deletions of ~ 100 bp are frequently reported in plants (Brooks et al., 2014; Kapusi et al., 2017; Nekrasov et al., 2017; Ordon et al., 2017). Dual-targeting based on the paired use of gRNAs, was also effective in deleting larger fragments (10–12 kb) as reported in rice and Arabidopsis (Durr et al., 2018; Wang et al., 2017) and even larger fragments of 170–245 kb were deleted by multiplex targeting in rice (Zhou et al., 2014). However, compared to point mutagenesis (effect of a single gRNA), genomic deletions (effect of paired gRNAs) consistently occurred at much lower rate even when two or more gRNAs of equal efficiencies were used (Minkenberg et al. 2017; Tian et al., 2017).

Thanks to the modular design for GoldenBraid constructs (Sarrion-Perdigones et al., 2013; Vazquez-Vilar et al., 2015; Vazquez-Vilar et al., 2016), we generated three final vectors with all needed transcriptional units (TUs). This modular system improved the versatility and combinatorial potentialty of technique. The TUs were combined in a binary vector following a double-loop iterative cloning strategy that allows the assembly of increasingly complex multigenic modules. The vectors containing the hCas9 CDS were placed under the control of optimal promoters for tomato species 35S (GB0639). Similarly, the gRNA multiplexed transcriptional unit (TU) were placed under the control of the AtU6-26 promoter (GB1001). Specific resistance gene were also introduced into each vector for in vitro selection, as required by crop transformation protocols. Three final constructs, containing single or double guides, were obtained as shown in figure 5. The correct insertion of

constructs in *A.tumefaciens* was verified by PCR amplification with specific primers (see supplementary table 2) using an aliquot of bacterial culture transformed with the relative expression construct. All constructs were also validated by sequencing.

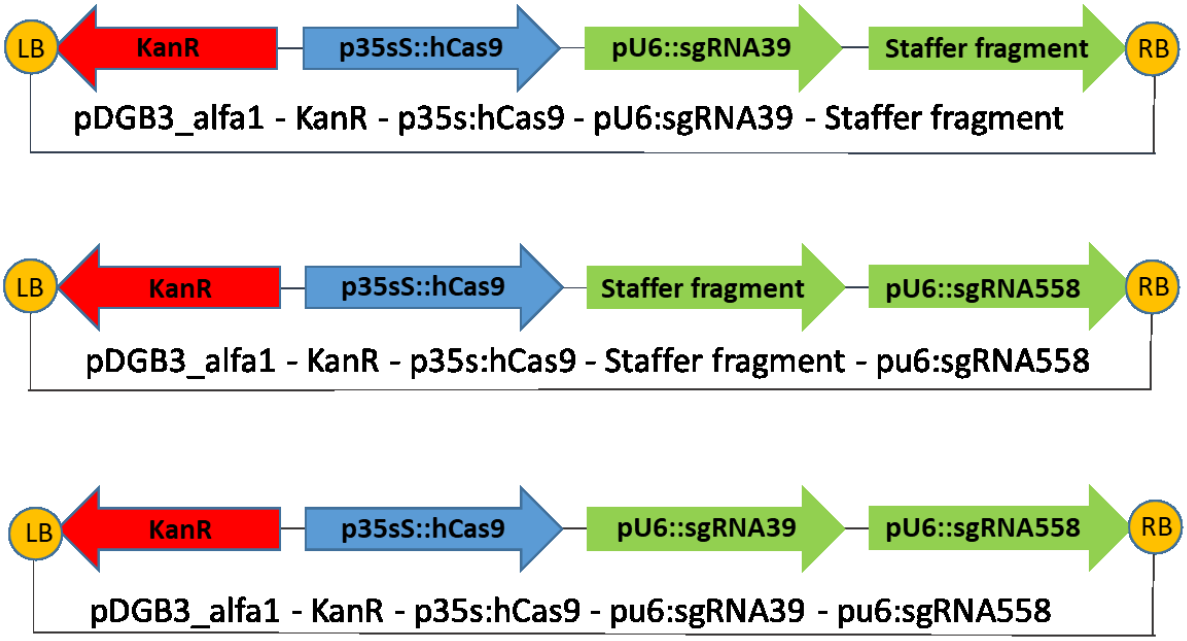
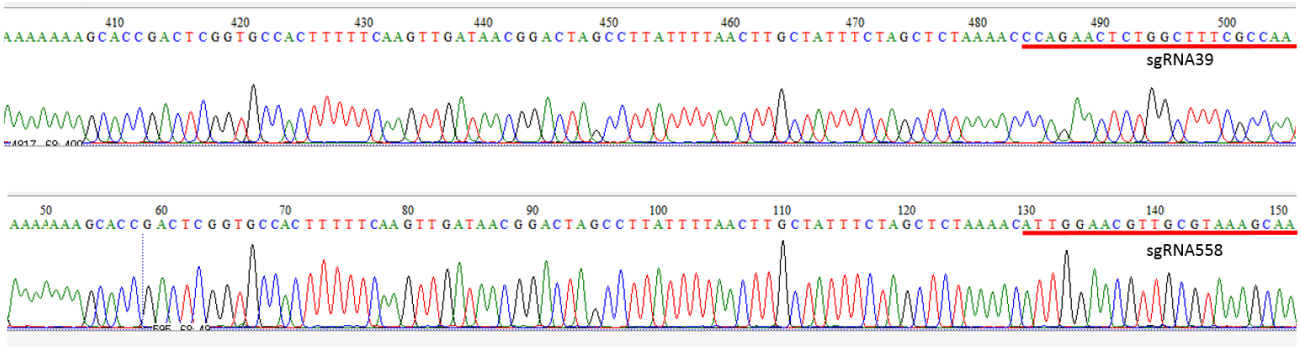



**Figure 5. Description of transcriptional units (arrows) assembled in the final vectors generated for plant transformation using the modular system GoldenBraid.**

The final GoldenBraid constructs were introduced into tomato plants as described in the materials and method section (4.2.4). *In vitro* selection of callus from tomato leaves allowed the regeneration of kanamycin resistance transgenic tomato lines only for the construct containing the dual-targeting by CRISPR/Cas9, based on the paired use of gRNAs (sgRNA39 and sgRNA558). For this construct, the PCR amplification confirmed the correct insertion in *A.tumefaciens,* (figure 6). At moment transformations experiments are ongoing and the CRISPR editing efficiency will be confirmed by the sequencing of targeted amplicons on primary transformant plants obtained.

a)


b)




c)

**Figure 6. Key steps in the construct validation and plant transformation. a) PCR amplification using an aliquot of 4 bacterial cultures transformed with the relative expression construct. The *nptII* gene (kanamycin resistance determinant) was amplified (753 bp). b) Sanger sequencing chromatograms confirm the final construct. The sequence obtained using a reverse primer (RB-TDNA Rv-see supplementary table 2) was displayed. The red lines highlight the different portions related to the two gRNAs (sgRNA39 and sgRNA558). c) Different stages of *Agrobacterium tumefaciens*-mediated transformation are shown.**

## 4.4 SUPPLEMENTARY

a)

| Bulge Type | Target | Chromosome | Position | Direction | Mismatches | Bulge Size |
|---|---|---|---|---|---|---|
| RNA | crRNA: GGCGAAAGCCAGAGTTCTGGNGG<br>DNA: tGCGAAAGaCAGAGT--TGGAGG | chr1 | 78840016 | + | 2 | 2 |
| RNA | crRNA: GGCGAAAGCCAGAGTTCTGGNGG<br>DNA: tGCGAAAGaCAGAGTT--GGAGG | chr1 | 78840016 | + | 2 | 2 |
| RNA | crRNA: GGCGAAAGCCAGAGTTCTGGNGG<br>DNA: aG-GAAAGCCAGAtTTCTGGTGG | chr10 | 55997973 | - | 2 | 1 |

b)

| Bulge Type | Target | Chromosome | Position | Direction | Mismatches | Bulge Size |
|---|---|---|---|---|---|---|
| RNA | crRNA: GCTTTACGCAACGTTCCAATNGG<br>DNA: GCTTTAgGgA--GTTCCAATGGG | chr7 | 223189 | - | 2 | 2 |
| RNA | crRNA: GCTTTACGCAACGTTCCAATNGG<br>DNA: GtTTTAtGCAAC--TCCAATTGG | chr1 | 16664391 | - | 2 | 2 |

**Supplementary table 1. Possible off targets identified for sgRNA39(a) and sgRNA558 (b).**

| Primer list |
|---|
| NPTII_F: 5’ ACGTGCTATTCGGCTATGACTGGG 3’ |
| NPTII_R: 5’TCAGAAGAACTCGTCAAGAAGGCG 3’ |
| LB-TDNA_F: 5’ TGGCAGGATATATTGTGGTG 3’ |
| RB-TDNA_R: 5’ TTACCCGCCAATATATCCTG 3’ |
| DOF11_558_F: 5' ATTGCTTTACGCAACGTTCCAAT 3' |
| DOF11_558_R: 5' AAACATTGGAACGTTGCGTAAAG 3’ |
| DOF11_39_F: 5' ATTGGCGAAAGCCAGAGTTCTGG 3' |
| DOF11_39_R: 5' AAACCCAGAACTCTGGCTTTCGC 3' |

**Supplementary table 2. List of primers used in this study.**

# CHAPTER V

## CONCLUSIONS

In next years, several agricultural regions will be more vulnerable to climate change as result of challenges such as temperature extremes, higher hot season average temperature, persistent drought. Tomato *(Solanum lycopersicum)* is one of the most important worldwide crop in terms of production despite being sensitive to high-temperature. Developing tomato cultivars tolerant to heat stress may be a valuable strategy to cope with climate changes. In this thesis, different strategies were integrated to promote the identification of traits involved in heat stress tolerance and the selection of elite lines tolerant to heat stress:

- ✓ The agronomic evaluation of a JAGF4 segregating population, under heat stress conditions, allowed us to identify the best performing lines as well as the worsts by the means of a PCA analysis. Most of the variation among the genotypes was explained by YP, TFN, FS. By contrast, SSC, IN and FRL account for the 13% of estimated variance. Fruit set and related flowering traits high impact total yield and SCC. A trade-off between these two parameters during the selection was imposed by their negative correlation.

- ✓ The characterization for sub-traits related to flowering of extreme genotypes confirmed the tolerant genotypes responded better to heat stress than the susceptible. Moreover, a high positive correlation among pollen viability (PV), total number of flowers (TNF) per plant and total number of inflorescences (TNI) and YP was found.

- ✓ The genotypes, showing highest PV, TNF and TNI values, were also YP performers, supporting the finding that the assessment of complex fruit setting and yield traits can be assisted through sub-traits (such as pollen viability). In addition, the heat stress test performed on limited area of cultivar *Monymaker* suggests that heat stress response is a local process.

- ✓ The 100 JAGF4 individuals previously phenotyped were also genotyped through GBS (genotyping by sequencing) approach. Optimized genomic prediction models for YP and

SSC, were developed testing several critical parameters, including training population size and composition and marker density and quality.

- ✓ GS results highlight that GS is a promising strategy to accelerate breeding for heat tolerance in tomato since our models were able to predict the GEBVs (Genomic estimated breeding values) in next round of selection (F5- F6). Moreover, the F6 offspring, grown under field heat stress condition, validated the prediction accuracy of the model, confirming the performance of the 9 top lines.

- ✓ Furthermore, the meaningful SNPs (calculated by the model) were used to perform a QTL analysis to better understand the genetic architecture of heat tolerant traits in tomato. Five significant QTLs were detected for YP under heat stress, on chromosome 5, 6, 8, 9 and 11 (qYP1, qYP2, qYP3, qYP4, qYP5) and 1 significant QTL for SSC, co-localizing with qYP2, on chromosome 6 (qSSC1). Yield-SCC contrasting pattern may be solved by fine genomic selection of such region.

- ✓ A variant annotation was performed on genes located in regions underlying identified QTL, to predict their effect on the protein function. Two suitable candidate genes were detected (*SlMYB59* and SKP1-like protein) that can be further functionally exploited. If the candidate genes will confirm to be involved in the heat-tolerance response could be then transferred to cultivated tomatoes to improve performances under high temperatures.

- ✓ A tomato genome scan and a phylogenetic analysis of Dof proteins family, including Arabidopsis members, allowed us to identify a putative transcription factor (SlDof) involved in the abiotic stress tolerance. The SlDof11 gene was choose as suitable target for performing a CRISPR/Cas9 experiment to validate its putative involvement in heat tolerance response.

# REFERENCES

Andolfo, G., Ferriello, F., Tardella, L., Ferrarini, A., Sigillo, L., Frusciante, L., Ercolano, M.R., 2014. Tomato genome-wide transcriptional responses to *Fusarium* wilt and Tomato Mosaic Virus. PLoS One 9(5), e94963.

Apel, K., Hirt, H., 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu Rev Plant Biol 55, 373-399.

Araus, J.L., Kefauver, S.C., Zaman-Allah, M., Olsen, M.S., Cairns, J.E., 2018. Translating High-Throughput Phenotyping into Genetic Gain. Trends Plant Sci 23(5), 451-466.

Asseng, S., Foster, I., Turner, N.C., 2011. The impact of temperature variability on wheat yields. Glob. Chang. Biol. 17, 997–1012. doi:10.1111/j.1365-2486.2010.02262.x

Ayenan, M.A.T., Agyemang, Danquah A., Hanson, P., Ampomah-Dwamena, C., Sodedji, F.A.K., As ante, I.K, Danquah, E.Y. 2019. Accelerating Breeding for Heat Tolerance in Tomato (*Solanum lycopersicum L.*): An Integrated Approach. Agronomy 9, 720; doi:10.3390/agronomy9110720

Bagha, S., 2014. The impact of chronic high temperatures on anther and pollen development in cultivated *Oryza* species. University of Toronto.

Bailey-Serres, J., Lee, S.C., Brinton, E., 2012. Waterproofing crops: effective flooding survival strategies. Plant Physiol 160(4), 1698-1709.

Balasubramanian, S., Weigel, D., 2006. Temperature Induced Flowering in *Arabidopsis thaliana*. Plant Signal Behav 1(5), 227-228.

BarnabÃ¡s, B., JÃ¤ger, K., FehÃ©r, A., 2008. The effect of drought and heat stress on reproductive processes in cereals. Plant Cell Environ 31(1), 11-38.

Bassi, F.M., Bentley, A.R., Charmet, G., Ortiz, R., Crossa, J., 2016. Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). Plant Sci 242, 23-36.

Bernardo, R., Yu, Y. 2007. Prospects for genome-wide selection for quantitative traits in maize. Crop Sci 47: 1082–1090.

Bita, C.E., Gerats, T., 2013. Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. Front Plant Sci 4, 273.

Bita, C.E., Zenoni, S., Vriezen, W.H., Mariani, C., Pezzotti, M., Gerats, T., 2011b. Temperature stress differentially modulates transcription in meiotic anthers of heat-tolerant and heat-sensitive tomato plants. BMC Genomics 12, 384.

Bo, C., Chen, H., Luo, G., Li, W., Zhang, X., Ma, Q., Cheng, B., Cai, R., 2020. Maize *WRKY114* gene negatively regulates salt-stress tolerance in transgenic rice. Plant Cell Rep 39(1), 135-148.

Bostock, R.M., Pye, M.F., Roubtsova, T.V., 2014a. Predisposition in plant disease: exploiting the nexus in abiotic and biotic stress perception and response. Annu Rev Phytopathol 52, 517-549.

Broman, K. W., Wu, H., Sen, Ś., and Churchill, G. A. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19, 889–890. doi:10.1093/bioinformatics/btg112.

Broman, K.W., Wu, H., Sen, S., Churchill, G.A., 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19(7), 889-890.

Brooks, C., Nekrasov, V., Lippman, Z.B., Van Eck, J., 2014. Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. Plant Physiol 166(3), 1292-1297.

Cai, X., Zhang, Y., Zhang, C., Zhang, T., Hu, T., Ye, J., Zhang, J., Wang, T., Li, H., Ye, Z., 2013a. Genome-wide analysis of plant-specific Dof transcription factor family in tomato. J Integr Plant Biol 55(6), 552-566.

Cao, D., Froehlich, J.E., Zhang, H., Cheng, C.L., 2003. The chlorate-resistant and photomorphogenesis-defective mutant *cr88* encodes a chloroplast-targeted HSP90. Plant J 33(1), 107-118.

Capuozzo, C., Formisano, G., Iovieno, P., Andolfo, G., Tomassoli, L., Barbella, M.M, et al. 2017. Inheritance analysis and identification of SNP markers associated with ZYMV resistance in *Cucurbita pepo*. Mol. Breeding. 37(8), 1-12. doi: 10.1007/s11032-017-0698-5

Casaretto, J.A., El-Kereamy, A., Zeng, B., Stiegelmeyer, S.M., Chen, X., Bi, Y.M., Rothstein, S.J., 2016. Expression of OsMYB55 in maize activates stress-responsive genes and enhances heat and drought tolerance. BMC Genomics 17, 312.

Castorina, G., Fox, S., Tonelli, C., Galbiati, M., Conti, L., 2016. A novel role for STOMATAL CARPENTER 1 in stomata patterning. BMC Plant Biol 16(1), 172.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol Ecol 22(11), 3124-3140.

Cericola, F., Jahoor, A., Orabi, J., Andersen, J.R., Janss, L.L., Jensen, J., 2017. Optimizing Training Population Size and Genotyping Strategy for Genomic Prediction Using Association Study Results

and Pedigree Information. A Case of Study in Advanced Wheat Breeding Lines. PLoS One 12(1), e0169606.

Chandrasegaran, S., Carroll, D., 2016. Origins of Programmable Nucleases for Genome Engineering. J Mol Biol 428(5 Pt B), 963-989.

Chen, J.Q., Meng, X.P., Zhang, Y., Xia, M., Wang, X.P., 2008. Over-expression of *OsDREB* genes lead to enhanced drought tolerance in rice. Biotechnol Lett 30(12), 2191-2198.

Collard, B.C., Mackill, D.J., 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363(1491), 557-572.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013a. Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121), 819-823.

Corrales, A.R., Carrillo, L., Lasierra, P., Nebauer, S.G., Dominguez-Figueroa, J., Renau-Morata, B., Pollmann, S., Granell, A., Molina, R.V., Vicente-Carbajosa, J., Medina, J., 2017. Multifaceted role of cycling DOF factor 3 (CDF3) in the regulation of flowering time and abiotic stress responses in *Arabidopsis*. Plant Cell Environ 40(5), 748-764.

Corrales, A.R., Nebauer, S.G., Carrillo, L., FernÃ¡ndez-Nohales, P., MarquÃ©s, J., Renau-Morata, B., Granell, A., Pollmann, S., Vicente-Carbajosa, J., Molina, R.V., Medina, J., 2014a. Characterization of tomato Cycling Dof Factors reveals conserved and new functions in the control of flowering time and abiotic stress responses. J Exp Bot 65(4), 995-1012.

Cramer, W., Guiot, J., Fader, M., Garrabou, J., Gattuso, J.P., Iglesias, A., Lange, M.A., Lionello, P., Llasat, M.C., Paz, S., et al. 2018. Climate change and interconnected risks to sustainable development in the Mediterranean. Nat. Clim. Chang. 8, 972–980.

Crossa, J., Campos, G.e.L., PÃ©rez, P., Gianola, D., BurgueÃ±o, J., Araus, J.L., Makumbi, D., Singh, R.P., Dreisigacker, S., Yan, J., Arief, V., Banziger, M., Braun, H.J., 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186(2), 713-724.

Crossa, J., PÃ©rez-RodrÃ-guez, P., Cuevas, J., Montesinos-LÃ³pez, O., JarquÃ-n, D., de Los Campos, G., BurgueÃ±o, J., GonzÃ¡lez-Camacho, J.M., PÃ©rez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., Varshney, R.K., 2017b. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends Plant Sci 22(11), 961-975.

Cui, Y., Li, R., Li, G., Zhang, F., Zhu, T., Zhang, Q., Ali, J., Li, Z., Xu, S., 2020. Hybrid breeding of rice via genomic selection. Plant Biotechnol J 18(1), 57-67.

D'Agostino, N., Tripodi, P. 2017. NGS-based genotyping, high-throughput phenotyping and genome-wide association studies laid the foundations for next-generation breeding in horticultural crops. Diversity 9, 38.

Dane, F., Hunter, A.G., Chambliss, O.L., 1991. Fruit Set, Pollen Fertility, and Combining Ability of Selected Tomato Genotypes under High-Temperature Field Conditions. J. Am. Soc. Hort. Sci. 116, 906–910.

Daniel, I.O., Atinsola, K.O., Ajala, M.O. and Popoola, A.R. 2017. Phenotyping a Tomato Breeding Population by Manual Field Evaluation and Digital Imaging Analysis. International Journal of Plant Breeding and Genetics, 11, 19-24.

De Storme, N., Geelen, D., 2014. The impact of environmental stress on male reproductive development in plants: biological processes and molecular mechanisms. Plant Cell Environ 37(1), 1-18.

Dempster, A. P., Laird, N. M., and Rubin, D. B. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. J. R. Stat. Soc. 39, 1–38.

Djanaguiraman, M., Prasad, P.V., Seppanen, M., 2010. Selenium protects sorghum leaves from oxidative damage under high temperature stress by enhancing antioxidant defense system. Plant Physiol Biochem 48(12), 999-1007.

Dreccer, M.F., Molero, G., Rivera-Amado, C., John-Bejai, C., Wilson, Z., 2019. Yielding to the image: How phenotyping reproductive growth can assist crop improvement and production. Plant Sci 282, 73-82.

Driedonks, N., Rieu, I., Vriezen, W.H., 2016. Breeding for plant heat tolerance at vegetative and reproductive stages. Plant Reprod 29(1-2), 67-79.

Driedonks, N., Wolters-Arts, M., Huber, H., de Boer, G.J., Vriezen, W., Mariani, C., Rieu, I., 2018. Exploring the natural variation for reproductive thermotolerance in wild tomato species. Euphytica 214,67.

Driedonks, N., Xu, J., Peters, J.L., Park, S., Rieu, I., 2015. Multi-Level Interactions Between Heat Shock Factors, Heat Shock Proteins, and the Redox System Regulate Acclimation to Heat. Front Plant Sci 6, 999.

Du, X.Q., Wang, F.L., Li, H., Jing, S., Yu, M., Li, J., Wu, W.H., Kudla, J., Wang, Y., 2019. The Transcription Factor MYB59 Regulates K. Plant Cell 31(3), 699-714.

Duan, D., Fischer, S., Merz, P., Bogs, J., Riemann, M., Nick, P., 2016. An ancestral allele of grapevine transcription factor MYB14 promotes plant defence. J Exp Bot 67(6), 1795-1804.

Durr, J., Papareddy, R., Nakajima, K., Gutierrez-Marcos, J., 2018. Highly efficient heritable targeted deletions of gene clusters and non-coding regulatory regions in Arabidopsis using CRISPR/Cas9. Sci Rep 8(1), 4443.

Edwards, S.M., Buntjer, J.B., Jackson, R., Bentley, A.R., Lage, J., Byrne, E., Burt, C., Jack, P., Berry, S., Flatman, E., Poupard, B., Smith, S., Hayes, C., Gaynor, R.C., Gorjanc, G., Howell, P., Ober, E., Mackay, I.J., Hickey, J.M., 2019. The effects of training population design on genomic prediction accuracy in wheat. Theor Appl Genet 132(7), 1943-1952.

Endelman, J.B., 2011. New algorithm improves fine structure of the barley consensus SNP map. BMC Genomics 12, 407.

Ewas, M., Khames, E., Ziaf, K., Shahzad, R., Nishawy, E., Ali, F., Subthain, H., Amar, M.H., Ayaad, M., Ghaly, O., Luo, J., 2017. The Tomato DOF Daily Fluctuations 1, TDDF1 acts as flowering accelerator and protector against various stresses. Sci Rep 7(1), 10299.

Fahad, S., Bajwa, A.A., Nazir, U., Anjum, S.A., Farooq, A., Zohaib, A., Sadia, S., Nasim, W., Adkins, S., Saud, S., Ihsan, M.Z., Alharby, H., Wu, C., Wang, D., Huang, J., 2017a. Crop Production under Drought and Heat Stress: Plant Responses and Management Options. Front Plant Sci 8, 1147.

Fahad, S., Hussain, S., Saud, S., Hassan, S., Ihsan, Z., Shah, A.N., Wu, C., Yousaf, M., Nasim, W., Alharby, H., Alghabari, F., Huang, J., 2016. Exogenously Applied Plant Growth Regulators Enhance the Morpho-Physiological Growth and Yield of Rice under High Temperature. Front Plant Sci 7, 1250.

Fasani, E., DalCorso, G., Costa, A., Zenoni, S., Furini, A., 2019. The Arabidopsis thaliana transcription factor MYB59 regulates calcium signalling during plant growth and stress response. Plant Mol Biol 99(6), 517-534.

Feng, B.H., Han, Y.C., Xiao, Y.Y., Kuang, J.F., Fan, Z.Q., Chen, J.Y., Lu, W.J., 2016. The banana fruit Dof transcription factor MaDof23 acts as a repressor and interacts with MaERF9 in regulating ripening-related genes. J Exp Bot 67(8), 2263-2275.

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39(Web Server issue), W29-37.

Firon, N., Pressman, E., Meir, S., Khoury, R., Altahan, L., 2012. Ethylene is involved in maintaining tomato (*Solanum lycopersicum*) pollen quality under heat-stress conditions. AoB Plants 2012, pls024.

Foolad, M.R., 2007. Genome mapping and molecular breeding of tomato. Int J Plant Genomics 2007, 64358.

Fortes, A.M., Granell, A., Pezzotti, M., Bouzayen, M., 2017. Editorial: Molecular and Metabolic Mechanisms Associated with Fleshy Fruit Quality. Front Plant Sci 8, 1236.

Gerszberg, A., Hnatuszko-Konka, K. 2017. Tomato tolerance to abiotic stress: a review of most often engineered target sequences. Plant Growth Regul. 83,175–198.

Giuliano, C.J., Lin, A., Girish, V., Sheltzer, J.M., 2019. Generating Single Cell-Derived Knockout Clones in Mammalian Cells with CRISPR/Cas9. Curr Protoc Mol Biol 128(1), e100.

Goddard, M., 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136(2), 245-257.

Goddard, M.E., Hayes, B.J., 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet 10(6), 381-391.

Golam, F., Prodhan, Z.H., Nezhadahmadi, A., Rahman, M., 2012. Heat tolerance in tomato. Life Sci. J. 9, 1936–1950.

Grilli, G.V.G.; Braz, L.T.; Lemos, E.G.M., 2007. QTL identification for tolerance to fruit set in tomato by FAFLP markers. Crop Breed. Appl. Biotechnol. 7, 234–241.

Gupta, S., Malviya, N., Kushwaha, H., Nasim, J., Bisht, N.C., Singh, V.K., Yadav, D., 2015. Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. Planta 241(3), 549-562.

Hanjing, L., 2015. Heat stress transcription factors in anthers - their role in protecting tapetal cells and pollen from heat, PhD thesis, Radboud University, Nijmegen (NL).

Hasanuzzaman, M., Nahar, K., Alam, M.M., Roychowdhury, R., Fujita, M., 2013. Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. Int J Mol Sci 14(5), 9643-9684.

Hedhly, A., Hormaza, J.I., Herrero, M., 2009a. Global warming and sexual plant reproduction. Trends Plant Sci 14(1), 30-36.

Heffner, E.L., Sorrells, M.E., and Jannink, J. 2009. Genomic Selection for Crop Improvement. Crop. Sci. 49(1), 1–12. doi: 10.2135/cropsci2008.08.0512.

Heidmann, I., Schade-Kampmann, G., Lambalk, J., Ottiger, M., Di Berardino, M., 2016. Impedance Flow Cytometry: A Novel Technique in Pollen Analysis. PLoS One 11(11), e0165531.

Herrero, M., 2003. Male and female synchrony and the regulation of mating in flowering plants. Philos Trans R Soc Lond B Biol Sci 358(1434), 1019-1024.

Hickman, R., Van Verk, M.C., Van Dijken, A.J.H., Mendes, M.P., Vroegop-Vos, I.A., Caarls, L., Steenbergen, M., Van der Nagel, I., Wesselink, G.J., Jironkin, A., Talbot, A., Rhodes, J., De Vries, M., Schuurink, R.C., Denby, K., Pieterse, C.M.J., Van Wees, S.C.M., 2017. Architecture and Dynamics of the Jasmonic Acid Gene Regulatory Network. Plant Cell 29(9), 2086-2105.

Hu, H., Dai, M., Yao, J., Xiao, B., Li, X., Zhang, Q., Xiong, L., 2006a. Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. Proc Natl Acad Sci U S A 103(35), 12987-12992.

Huang, J., Sun, S.J., Xu, D.Q., Yang, X., Bao, Y.M., Wang, Z.F., Tang, H.J., Zhang, H., 2009. Increased tolerance of rice to cold, drought and oxidative stresses mediated by the overexpression of a gene that encodes the zinc finger protein ZFP245. Biochem Biophys Res Commun 389(3), 556-561.

Huang, L., Wang, Y., Wang, W., Zhao, X., Qin, Q., Sun, F., Hu, F., Zhao, Y., Li, Z., Fu, B., 2018. Characterization of Transcription Factor Gene. Front Plant Sci 9, 94.

Iba, K., 2002. Acclimative response to temperature stress in higher plants: approaches of gene engineering for temperature tolerance. Annu Rev Plant Biol 53, 225-245.

Jha, U.C., Bohra, A., Singh, N.P., 2014. Heat stress in crop plants: Its nature, impacts and integrated breeding strategies to improve heat tolerance. Plant Breed. 133, 679–701.

Jin, C., Li, K.Q., Xu, X.Y., Zhang, H.P., Chen, H.X., Chen, Y.H., Hao, J., Wang, Y., Huang, X.S., Zhang, S.L., 2017a. A Novel NAC Transcription Factor,. Front Plant Sci 8, 1049.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096), 816-821.

Jones, P., Binns D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. 2014. InterProScan 5: genome-scale protein function classification, Bioinformatics, 30, 1236-1240.

Kakani, V.G., Reddy, K.R., Koti, S., Wallace, T.P., Prasad, P.V., Reddy, V.R., Zhao, D., 2005. Differences in in vitro pollen germination and pollen tube growth of cotton cultivars in response to high temperature. Ann Bot 96(1), 59-67.

Kanbar, A., Kondo, K., and Shashidhar, H.E., 2011. Comparative efficiency of pedigree, modified bulk and single seed descent breeding methods of selection for developing high-yielding lines in rice (*Oryza sativa L.*) under aerobic condition. Electron. J. Plant Breed. 2, 184–193.

Kang, W.H., Kim, S., Lee, H.A., Choi, D., Yeom, S.I., 2016. Genome-wide analysis of Dof transcription factors reveals functional characteristics during development and response to biotic stresses in pepper. Sci Rep 6, 33332.

Kapusi, E., Corcuera-GÃ³mez, M., Melnik, S., Stoger, E., 2017. Heritable Genomic Fragment Deletions and Small Indels in the Putative ENGase Gene Induced by CRISPR/Cas9 in Barley. Front Plant Sci 8, 540.

Kartikeya, S., Sunil, K., Surender, K., Pravin, P., Vaishampayan, A., 2012. Screening of tomato genotypes for reproductive characters under high temperature stress conditions. Sabrao J. Breed. Genet. 44 (2):263276

Kim, N.Y., Jang, Y.J., Park, O.K., 2018a. AP2/ERF Family Transcription Factors ORA59 and RAP2.3 Interact in the Nucleus and Function Together in Ethylene Responses. Front Plant Sci 9, 1675.

Klap, C., Yeshayahou, E., Bolger, A.M., Arazi, T., Gupta, S.K., Shabtai, S., Usadel, B., Salts, Y., Barg, R., 2017a. Tomato facultative parthenocarpy results from SlAGAMOUS-LIKE 6 loss of function. Plant Biotechnol J 15(5), 634-647.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009a. Circos: an information aesthetic for comparative genomics. Genome Res 19(9), 1639-1645.

Kumar, S., KrabberÃ¸d, A.K., Neumann, R.S., Michalickova, K., Zhao, S., Zhang, X., Shalchian-Tabrizi, K., 2015. BIR Pipeline for Preparation of Phylogenomic Data. Evol Bioinform Online 11, 79-83.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23(21), 2947-2948.

Lee, S.J., Kang, J.Y., Park, H.J., Kim, M.D., Bae, M.S., Choi, H.I., Kim, S.Y., 2010a. DREB2C interacts with ABF2, a bZIP protein regulating abscisic acid-responsive gene expression, and its overexpression affects abscisic acid sensitivity. Plant Physiol 153(2), 716-727.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14), 1754-1760.

Li, Z., Li, L., Zhou, K., Zhang, Y., Han, X., Din, Y., Ge, X., Qin, W., Wang, P., Li, F., Ma, Z., Yang, Z., 2019. Acts as a Negative Regulator in Both Transgenic. Front Genet 10, 392.

Liabeuf, D., Sim, S.C., Francis, D.M., 2018. Comparison of marker-based genomic estimated breeding values and phenotypic evaluation for selection of bacterial spot resistance in tomato. Phytopathology, 108: 392-401.

Lin, F., Wan, S.Q., Cheng, L.G., LÃ¼, H.Y., Li, G.J., Zhang, Y.M., 2006. [A simulated study on mapping QTL in a segregating sub-population]. Yi Chuan 28(11), 1407-1410.

Liu, F., Ni, W., Griffith, M., Huang, Z., Chang, C., Peng, W., Ma, H., and Xie, D., 2004. The *ASK1* and *ASK2* genes are essential for *Arabidopsis* early development. Plant Cell 16: 5–20.

Liu, X., Liu, Z., Hao, Z., Chen, G., Qi, K., Zhang, H., Jiao, H., Wu, X., Zhang, S., Wu, J., Wang, P., 2020. Characterization of Dof family in *Pyrus bretschneideri* and role of PbDof9.2 in flowering time regulation. Genomics 112(1), 712-720.

Lorenz, A.J., Chao, S., Asoro, F.G., Heffner, E.L., Hayashi, T., Iwata, H., et al. 2011. Genomic Selection in Plant Breeding. Knowledge and Prospects. Advances in Agronomy. 110, 77-123.

Lorenzana, R.E., and Bernardo, R., 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. 120, 151–161. doi: 10.1007/s00122-009-1166-3.

Lorrai, R., Gandolfi, F., Boccaccini, A., Ruta, V., Possenti, M., Tramontano, A., Costantino, P., Lepore, R., Vittorioso, P., 2018a. Genome-wide RNA-seq analysis indicates that the DAG1 transcription factor promotes hypocotyl elongation acting on ABA, ethylene and auxin signaling. Sci Rep 8(1), 15895.

Lu, G., Gao, C., Zheng, X., Han, B., 2009. Identification of OsbZIP72 as a positive regulator of ABA response and drought tolerance in rice. Planta 229(3), 605-615.

Muller, F., Xu, J., Kristensen, L., Wolters-Arts, M., de Groot, P.F., Jansma, S.Y., Mariani, C., Park, S., Rieu, I., 2016. High-Temperature-Induced Defects in Tomato (Solanum lycopersicum) Anther and Pollen Development Are Associated with Reduced Expression of B-Class Floral Patterning Genes. PLoS One 11(12), e0167614.

Mali, P., Esvelt, K.M., Church, G.M., 2013. Cas9 as a versatile tool for engineering biology. Nat Methods 10(10), 957-963.

Malzahn, A., Lowder, L., Qi, Y., 2017. Plant genome editing with TALEN and CRISPR. Cell Biosci 7, 21.

Matsui, T., Omasa, K., 2002. Rice (Oryza sativa L.) cultivars tolerant to high temperature at flowering: anther characteristics. Ann Bot 89(6), 683-687.

Mazzucato, A., Papa, R., Bitocchi, E., Mosconi, P., Nanni, L., Negri, V., Picarella, M.E., Siligato, F., Soressi, G.P., Tiranti, B., et al. 2008. Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum L.*) landraces. TAG Theor. Appl. Genet. 116, 657–669.

Mesihovic, A., Iannacone, R., Firon, N., Fragkostefanakis, S., 2016. Heat stress regimes for the investigation of pollen thermotolerance in crop plants. Plant Reprod 29(1-2), 93-105.

Meuwissen, T., Hayes, B., Goddard, M., 2013b. Accelerating improvement of livestock with genomic selection. Annu Rev Anim Biosci 1, 221-237.

Minkenberg, B., Wheatley, M., Yang, Y., 2017. CRISPR/Cas9-Enabled Multiplex Genome Editing and Its Application. Prog Mol Biol Transl Sci 149, 111-132.

Mittler, R., 2006. Abiotic stress, the field environment and stress combination. Trends Plant Sci 11(1), 15-19.

Mittler, R., Kim, Y., Song, L., Coutu, J., Coutu, A., Ciftci-Yilmaz, S., Lee, H., Stevenson, B., Zhu, J.K., 2006. Gain- and loss-of-function mutations in Zat10 enhance the tolerance of plants to abiotic stress. FEBS Lett 580(28-29), 6537-6542.

Morimoto, R.I., 1998. Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes Dev 12(24), 3788-3796.

Morrison, M.J. and D.W. Stewart, 2002. Heat stress during flowering in summer brassica. Crop Sci., 42: 797–803.

Mu, R.L., Cao, Y.R., Liu, Y.F., Lei, G., Zou, H.F., Liao, Y., Wang, H.W., Zhang, W.K., Ma, B., Du, J.Z., Yuan, M., Zhang, J.S., Chen, S.Y., 2009. An R2R3-type transcription factor gene AtMYB59 regulates root growth and cell cycle progression in Arabidopsis. Cell Res 19(11), 1291-1304.

Negi, J., Moriwaki, K., Konishi, M., Yokoyama, R., Nakano, T., Kusumi, K., Hashimoto-Sugimoto, M., Schroeder, J.I., Nishitani, K., Yanagisawa, S., Iba, K., 2013. A Dof transcription factor, SCAP1, is essential for the development of functional stomata in Arabidopsis. Curr Biol 23(6), 479-484.

Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D., Kamoun, S., 2017. Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. Sci Rep 7(1), 482.

Nguyen, T.V., Do, L.T.K., Somfai, T., Otoi, T., Taniguchi, M., Kikuchi, K., 2019a. Presence of chlorogenic acid during in vitro maturation protects porcine oocytes from the negative effects of heat stress. Anim Sci J 90(12), 1530-1536.

Nishida, S., Kakei, Y., Shimada, Y., Fujiwara, T., 2017. Genome-wide analysis of specific alterations in transcript structure and accumulation caused by nutrient deficiencies in Arabidopsis thaliana. Plant J 91(4), 741-753.

Noguero, M., Atif, R.M., Ochatt, S., Thompson, R.D., 2013. The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. Plant Sci 209, 32-45.

Oh, S., Park, S., Han, K.H., 2003. Transcriptional regulation of secondary growth in *Arabidopsis thaliana*. J Exp Bot 54(393), 2709-2722.

Ordon, J., Gantner, J., Kemna, J., Schwalgun, L., Reschke, M., Streubel, J., Boch, J., Stuttmann, J., 2017. Generation of chromosomal deletions in dicotyledonous plants employing a user-friendly genome editing toolkit. Plant J 89(1), 155-168.

Pachauri, R.K., Allen, M.R., Barros, V.R., Broome, J., Cramer, W., Christ, R., Church, J.A., Clarke, L., Dahe, Q., Dasgupta, P., others, 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

.

Panthee, D.P., Labate, J.A., McGrath, M.T., Breksa, A.P. III, Robertson, L.D. 2013. Genotype and environmental interaction for fruit quality traits in vintage tomato varieties. Euphytica 193:169–182.

Panthee, D.R., Piotrowski, A., Ibrahem, R., 2017. Mapping Quantitative Trait Loci (QTL) for Resistance to Late Blight in Tomato. Int J Mol Sci 18(7).

Parish, R.W., Phan, H.A., Iacuone, S., Li, S.F., 2012. Tapetal development and abiotic stress: a centre of vulnerability. Funct Plant Biol 39:553–559. doi:10.1071/FP12090.

Pasquali, G., Biricolti, S., Locatelli, F., Baldoni, E., Mattana, M., 2008. Osmyb4 expression improves adaptive responses to drought and cold stress in transgenic apples. Plant Cell Rep 27(10), 1677-1686.

Paul, J.W., Qi, Y., 2016. CRISPR/Cas9 for plant genome editing: accomplishments, problems and prospects. Plant Cell Rep 35(7), 1417-1427.

Paupiere, M.J., Muller, F., Li, H., Rieu, I., Tikunov, Y.M., Visser, R.G.F., Bovy, A.G., 2017. Untargeted metabolomic analysis of tomato pollen development and heat stress response. Plant Reprod 30(2), 81-94.

Peng, J., Qi, X., Chen, X., Li, N., Yu, J., 2017. ZmDof30 Negatively Regulates the Promoter Activity of the Pollen-Specific Gene. Front Plant Sci 8, 685.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7(5), e37135.

Prasad, P.V.V., Djanaguiraman, M., 2011. High night temperature decreases leaf photosynthesis and pollen function in grain sorghum. Functional Plant Biology 38 (12):993-1003.

Pressman, E., Peet, M.M., Pharr, D.M., 2002. The effect of heat stress on tomato pollen characteristics is associated with changes in carbohydrate concentration in the developing anthers. Ann Bot 90(5), 631-636.

Qin, H., Wang, J., Chen, X., Wang, F., Peng, P., Zhou, Y., Miao, Y., Zhang, Y., Gao, Y., Qi, Y., Zhou, J., Huang, R., 2019. Rice OsDOF15 contributes to ethylene-inhibited primary root elongation under salt stress. New Phytol 223(2), 798-813.

Rao, A.V., Agarwal, S., 2000. Role of antioxidant lycopene in cancer and heart disease. J Am Coll Nutr 19(5), 563-569.

Renau-Morata, B., Molina, R.V., Carrillo, L., Cebolla-Cornejo, J., Sánchez-Perales, M., Pollmann, S., Domínguez-Figueroa, J., Corrales, A.R., Flexas, J., Vicente-Carbajosa, J., et al. 2017. Ectopic expression of CDF3 genes in tomato enhances biomass production and yield under salinity stress conditions. Front. Plant Sci. 8, 660.

Robertsen, C.D., Hjotrtshøj, R.L., and Janss, L.L., 2019. Genomic Selection in Cereal Breeding. Agronomy. 9, 1-16.

Rodriguez, G.R., Moyseenko, J.B., Robbins, M.D., MorejÃ³n, N.H., Francis, D.M., van der Knaap, E., 2010. Tomato Analyzer: a useful software application to collect accurate and detailed morphological and colorimetric data from two-dimensional objects. J Vis Exp(37).

Rodriguez, G.R., MuÃ±os, S., Anderson, C., Sim, S.C., Michel, A., Causse, M., Gardener, B.B., Francis, D., van der Knaap, E., 2011. Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. Plant Physiol 156(1), 275-285.

Rojas-Gracia, P., Roque, E., Medina, M., López-Martín, M. J., Cañas, L. A., Beltrán, J. P., et al. 2019. The DOF transcription factor SlDOF10 regulates vascular tissue formation during ovary development in tomato. Front. Plant Sci. 10:216. doi: 10.3389/fpls.2019.00216.

Ruggieri, V., Calafiore, R., Schettini, C., Rigano, M.M., Olivieri, F., Frusciante, L., Barone, A., 2019. Exploiting Genetic and Genomic Resources to Enhance Heat-Tolerance in Tomatoes. Agronomy 9, 22.

Rymen, B., Kawamura, A., SchÃ¤fer, S., Breuer, C., Iwase, A., Shibata, M., Ikeda, M., Mitsuda, N., Koncz, C., Ohme-Takagi, M., Matsui, M., Sugimoto, K., 2017. ABA Suppresses Root Hair Growth via the OBP4 Transcriptional Regulator. Plant Physiol 173(3), 1750-1762.

Saha, S., Hossain, M., Rahman, M., Kuo, C., Abdullah, S., 2010. Effect of high temperature stress on the performance of twelve sweet pepper genotypes. Bangladesh J. Agric. Res. 35, 525–534.

Sakata, T., Oshino, T., Miura, S., Tomabechi, M., Tsunaga, Y., Higashitani, N., Miyazawa, Y., Takahashi, H., Watanabe, M., Higashitani, A., 2010. Auxins reverse plant male sterility caused by high temperatures. Proc Natl Acad Sci U S A 107(19), 8569-8574.

Santopolo, S., Boccaccini, A., Lorrai, R., Ruta, V., Capauto, D., Minutello, E., Serino, G., Costantino, P., Vittorioso, P., 2015. DOF AFFECTING GERMINATION 2 is a positive regulator of light-mediated seed germination and is repressed by DOF AFFECTING GERMINATION 1. BMC Plant Biol 15, 72.

Sarinelli, J.M., Murphy, J.P., Tyagi, P., Holland, J.B., Johnson, J.W., Mergoum, M., Mason, R.E., Babar, A., Harrison, S., Sutton, R., Griffey, C.A., Brown-Guedira, G., 2019a. Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. Theor Appl Genet 132(4), 1247-1261.

Sarrion-Perdigones, A., Falconi, E. E., Zandalinas, S. I., Juarez, P., Fernandez-del-Carmen, A., Granell, A., and Orzaez, D. 2011. GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. PLoS One 6, e21622.

Sato, S., Peet, M.M., Thomas, J.F., 2000. Physiological factors limit fruit set of tomato (*Lycopersicon esculentum Mill.*) under chronic, mild heat stress. Plant Cell Environ. 23, 719–726.

Sato, S., Peet, M.M., Thomas, J.F., 2002. Determining critical pre- and post-anthesis periods and physiological processes in *Lycopersicon esculentum Mill*. exposed to moderately elevated temperatures. J Exp Bot 53(371), 1187-1195.

Shaheen, M.R., Ayyub, C.M., Amjad, M., Waraich, E.A., 2016 Morpho-physiological evaluation of tomato genotypes under high temperature stress conditions. J. Sci. Food Agric. 2016, 96, 2698–2704.

Shi, F., Dong, Y., Zhang, Y., Yang, X., Qiu, D., 2017. Overexpression of the. Front Plant Sci 8, 970.

Shi, Z., Wang, F., Cui, Y., Liu, Z., Guo, X., Zhang, Y., Deng, Y., Zhao, H., Chen, Y., 2015. Heritable CRISPR/Cas9-mediated targeted integration in Xenopus tropicalis. FASEB J 29(12), 4914-4923.

Shivanna, K.R., Linskens, H.F., Cresti, M., 1991. Pollen viability and pollen vigor. Theor Appl Genet 81(1), 38-42.

Singh, U., Patel, P.K., Singh, A.K., Tiwari, V., Kumar, R., Rai, N., Bahadur, A., Tiwari, S.K., Singh, M., Singh, B., 2015. Screening of tomato genotypes under high temperature stress for reproductive traits. Veg. Sci. 42, 52–55.

Song, J., Carver, B.F., Powers, C., Yan, L., KlÃ¡pÅ¡tÄ›, J., El-Kassaby, Y.A., Chen, C., 2017. Practical application of genomic selection in a doubled-haploid winter wheat breeding program. Mol Breed 37(10), 117.

Sun, Y., Yu, D., 2015. Activated expression of AtWRKY53 negatively regulates drought tolerance by mediating stomatal movement. Plant Cell Rep 34(8), 1295-1306.

Swinnen, G., Jacobs, T., Pauwels, L., Goossens, A., 2020. CRISPR-Cas-Mediated Gene Knockout in Tomato. Methods Mol Biol 2083, 321-341.

Takeoka, Y., Hiroi, K., Kitano, H., and Wada, T. 1991. Pistil hyperplasia in rice spikelets as affected by heat stress. Sex. Plant Reprod. 4, 39–43. doi:10.1007/BF00194570.

Tayade, R., Nguyen, T., Oh, S.A., Hwang, Y.S., 2018. Effective Strategies for Enhancing Tolerance to High-Temperature Stress in Rice during the Reproductive and Ripening Stages. Plant Breed Biotechnol. 2018; 6(1): 1–18

Tigist, M., Workneh, T.S., Woldetsadik, K., 2013. Effects of variety on the quality of tomato stored under ambient conditions. J Food Sci Technol 50(3), 477-486.

Tonsor, S.J., Scott, C., Boumaza, I., Liss, T.R., Brodsky, J.L., Vierling, E., 2008. Heat shock protein 101 effects in A. thaliana: genetic variation, fitness and pleiotropy in controlled temperature conditions. Mol Ecol 17(6), 1614-1626.

Vazquez-Vilar, M., Bernabe-Orts, J.M., Fernandez-del-Carmen, A., Ziarsolo, P., Blanca, J., Granell, A. and Orzaez, D. 2016. A modular toolbox for gRNA–Cas9 genome engineering in plants based on the GoldenBraid standard. Plant Met. 12, 1–12.

Vazquez-Vilar, M., Sarrion-Perdigones, A., Ziarsolo, P., Blanca, J., Granell, A. and Orzaez, D. 2015. Software-assisted stacking of gene modules using GoldenBraid 2.0 DNA-assembly framework. Methods Mol. Biol. 1284, 399–420.

Venkatesh, J., Park, S.W., 2015. Genome-wide analysis and expression profiling of DNA-binding with one zinc finger (Dof) transcription factor family in potato. Plant Physiol Biochem 94, 73-85.

Volkov, R.A., Panchuk, I.I., Mullineaux, P.M., SchÃ¶ffl, F., 2006. Heat stress-induced H(2)O (2) is required for effective expression of heat shock genes in Arabidopsis. Plant Mol Biol 61(4-5), 733-746.

Wahid, A., 2007. Physiological implications of metabolite biosynthesis for net assimilation and heat-stress tolerance of sugarcane (Saccharum officinarum) sprouts. J Plant Res 120(2), 219-228.

Waltz, E., 2018. With a free pass, CRISPR-edited plants reach market in record time. Nat Biotechnol 36(1), 6-7.

Wang X., Xu Y., Hu Z., Xu C., 2018. Genomic selection methods for crop improvement: current status and prospects. Crop J. 6 330–340. 10.1016/j.cj.2018.03.001.

Wang, C., Wang, K., 2017. [Advances in CRISPR-Cas-mediated genome editing system in plants]. Sheng Wu Gong Cheng Xue Bao 33(10), 1712-1722.

Wang, F., Chen, H.W., Li, Q.T., Wei, W., Li, W., Zhang, W.K., Ma, B., Bi, Y.D., Lai, Y.C., Liu, X.L., Man, W.Q., Zhang, J.S., Chen, S.Y., 2015. GmWRKY27 interacts with GmMYB174 to reduce expression of GmNAC29 for stress tolerance in soybean plants. Plant J 83(2), 224-236.

Wang, L., Ran, L., Hou, Y., Tian, Q., Li, C., Liu, R., Fan, D., Luo, K., 2017. The transcription factor MYB115 contributes to the regulation of proanthocyanidin biosynthesis and enhances fungal resistance in poplar. New Phytol 215(1), 351-367.

Wang, P., Li, J., Gao, X., Zhang, D., Li, A., Liu, C., 2018. Genome-Wide Screening and Characterization of the. Int J Mol Sci 19(6).

Warland, J., McKeown, A.W., McDonald, M.R. Impact of high air temperatures on *Brassicacae* crops in southern Ontario. Can. J. Plant Sci. 2006, 86, 1209–1215.

Weerakoon, W.M.W., Maruyama, A., Ohba, K., 2008 Impact of Humidity on Temperature-Induced Grain Sterility in Rice (*Oryza sativa L*). J. Agron. Crop Sci. 194, 135–140.

Wei, H., Zhao, H., Su, T., Bausewein, A., Greiner, S., Harms, K., Rausch, T., 2017. Chicory R2R3-MYB transcription factors CiMYB5 and CiMYB3 regulate fructan 1-exohydrolase expression in response to abiotic stress and hormonal cues. J Exp Bot 68(15), 4323-4338.

Wei, Q., Wang, W., Hu, T., Hu, H., Mao, W., Zhu, Q., Bao, C., 2018. Genome-wide identification and characterization of Dof transcription factors in eggplant (PeerJ 6, e4481.

Wei, W., Liang, D.W., Bian, X.H., Shen, M., Xiao, J.H., Zhang, W.K., Ma, B., Lin, Q., Lv, J., Chen, X., Chen, S.Y., Zhang, J.S., 2019. GmWRKY54 improves drought tolerance through activating genes in abscisic acid and Ca. Plant J 100(2), 384-398.

Wen, C.L., Cheng, Q., Zhao, L., Mao, A., Yang, J., Yu, S., Weng, Y., Xu, Y., 2016. Identification and characterisation of Dof transcription factors in the cucumber genome. Sci Rep 6, 23072.

Wen, J., Jiang, F., Weng, Y., Sun, M., Shi, X., Zhou, Y., Yu, L., Wu, Z., 2019. Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. BMC Plant Biol. 19, 1–17.

Wheeler, T., von Braun, J., 2013a. Climate change impacts on global food security. Science 341(6145), 508-513.

.

Wu, J., Chen, L., Chen, M., Zhou, W., Dong, Q., Jiang, H., Cheng, B., 2019. The DOF-Domain Transcription Factor ZmDOF36 Positively Regulates Starch Synthesis in Transgenic Maize. Front Plant Sci 10, 465.

Wu, Q., Li, D., Liu, X., Zhao, X., Li, X., Li, S., Zhu, L., 2015. Overexpression of OsDof12 affects plant architecture in rice (*Oryza sativa L.*). Front Plant Sci 6, 833.

Wu, Q., Liu, X., Yin, D., Yuan, H., Xie, Q., Zhao, X., Li, X., Zhu, L., Li, S., Li, D., 2017. Constitutive expression of OsDof4, encoding a C. BMC Plant Biol 17(1), 166.

Xu, J., Wolters-Arts, M., Mariani, C., Huber, H.; Rieu, I., 2017. Heat stress affects vegetative and reproductive performance and trait correlations in tomato (*Solanum lycopersicum*). Euphytica, 213, 1–12.

Xu, J., Driedonks, N., Rutten, M.J.M., Vriezen, W.H., de Boer, G.J., Rieu, I., 2017. Mapping quantitative trait loci for heat tolerance of reproductive traits in tomato (*Solanum lycopersicum*). Mol. Breed. New Strateg. Plant Improv. 37, 58.

Yamamoto, E., Matsunaga, H., Onogi, A., Ohyama, A., Miyatake, K., Yamaguchi, H., Nunome, T., Iwata, H., Fukuoka, H., 2017b. Efficiency of genomic selection for breeding population design and phenotype prediction in tomato. Heredity (Edinb) 118(2), 202-209.

Yan, W., Ye, S., Jin, Q., Zeng, L., Peng, Y., Yan, D., Yang, W., Yang, D., He, Z., Dong, Y., Zhang, X., 2010. Characterization and mapping of a novel mutant sms1 (senescence and male sterility 1) in rice. J Genet Genomics 37(1), 47-55.

Yanagisawa, M., Desyatova, A.S., Belteton, S.A., Mallery, E.L., Turner, J.A., Szymanski, D.B., 2015. Patterning mechanisms of cytoskeletal and cell wall systems during leaf trichome morphogenesis. Nat Plants 1, 15014.

Yanagisawa, S., 1998. [Dof proteins: involvement of transcription factors with a novel DNA-binding domain in tissue-specific and signal-responsive gene expression]. Seikagaku 70(4), 280-285.

Yanagisawa, S., 2001. The transcriptional activation domain of the plant-specific Dof1 factor functions in plant, animal, and yeast cells. Plant Cell Physiol 42(8), 813-822.

Yanagisawa, S., 2004. Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants. Plant Cell Physiol 45(4), 386-391.

Yanagisawa, S., and Izui, K. 1993. J. Biol. Chem.268: 16028–16036

Yanagisawa, S., and Sheen, J., 1998. Plant Cell10: 75–89.

Zafar, S.A., Patil, S., Uzair, M., Fang, J., Zhao, J., Yuan, S., Uzair, M., Luo, Q., Shi, J., Schreiber, L., Li, X., 2019. *DEGENERATED PANICLE AND PARTIAL STERILITY 1* (*DPS1*) encodes a CBS domain containing protein required for anther cuticle and panicle development in rice. New Phyt.

Zang, D., Wang, L., Zhang, Y., Zhao, H., Wang, Y., 2017. ThDof1.4 and ThZFP1 constitute a transcriptional regulatory cascade involved in salt or osmotic stress in Tamarix hispida. Plant Mol Biol 94(4-5), 495-507.

Zhang, A., Liu, Y., Wang, F., Li, T., Chen, Z., Kong, D., Bi, J., Zhang, F., Luo, X., Wang, J. 2019. Enhanced rice salinity tolerance via CRISPR/Cas9-targeted mutagenesis of the *OsRR22* gene. Mol. Breed. 39, 47.

Zhang, H., Yin, L., Wang, M., Yuan, X., Liu, X., 2019. Factors Affecting the Accuracy of Genomic Selection for Agricultural Economic Traits in Maize, Cattle, and Pig Populations. Front Genet 10, 189.

Zhou, H., Liu, B., Weeks, D.P., Spalding, M.H., Yang, B., 2014. Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. Nucleic Acids Res 42(17), 10903-10914.

Zhou, R., Yu, X., Kjær, K.H., Rosenqvist, E., Ottosen, C.O., Wu, Z., 2015. Screening and validation of tomato genotypes under heat stress using Fv/Fm to reveal the physiological mechanism of heat tolerance. Environ. Exp. Bot. 118, 1–11.

Zhu, L.P., Yu, Z., Zou, C.X., Li, Q.L., 2010. [Plant stress-inducible promoters and their function]. Yi Chuan 32(3), 229-234.

Zinn, K.E., Tunc-Ozdemir, M., Harper, J.F., 2010. Temperature stress and plant sexual reproduction: uncovering the weakest links. J Exp Bot 61(7), 1959-1968.

Zou, Z., Zhang, X., 2019. Genome-wide identification and comparative evolutionary analysis of the Dof transcription factor family in physic nut and castor bean. PeerJ 7, e6354.

Zou, Z., Zhu, J., Zhang, X., 2019a. Genome-wide identification and characterization of the Dof gene family in cassava (Manihot esculenta). Gene 687, 298-307.

# Acknowledgment

This work was funded from the European Union's Horizon 2020 research and innovation programme through the TomGEM project under grant agreement No 679796. In this contest I want sincerely thank Prof. Luigi Frusciante, Prof. Amalia Barone and Prof. Maria Raffaella Ercolano for having transmitted to me their knowledge and enthusiasm.

I sincerely thank my tutor, Prof. Maria Raffaella Ercolano for the opportunity she had given to me to deepen my knowledge on my research topic. I want also to thank her for having believed in me.

I want to thank all my collegous and in particular Prof. Antonio Di Matteo and Dott. Giuseppe Andolfo for their significant support during my PhD Program.

I wish to thank Prof. Ivo Rieu and Prof. Celestina Mariani for having hosted me as visiting PhD student at the Molecular Plant Physiology Department- Radboud University- Nijmegen (Netherlands) and for having wisely guided me in my research activities.